ULTRA-HIGH RESOLUTION DNA CRYSTALLOGRAPHY

A Thesis
Presented to
The Academic Faculty

by

Tatsuya Maehigashi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
May 2009

COPYRIGHT © 2009 TATSUYA MAEHIGASHI

ULTRA-HIGH RESOLUTION DNA CRYSTALLOGRAPHY

Approved by:

Dr. Loren D. Williams, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Dr. Lawrence A. Bottomley
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Adegboyega K. Oyelere
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: May 14, 2009

親不孝者

この言葉がたびたび脳裏をかすめる。日本を飛び出して何を好き勝手やってきたのだろう。この学位を受け取るとこが、なにかの節目になればと切に思う。

ACKNOWLEDGEMENTS

Firstly, I wish to acknowledge my major advisor and mentor, Professor Loren D. Williams for his support and continuing effort that made this research possible. I would also like to acknowledge Professor Nicholas V. Hud for his support and providing a great opportunity to work with intercalators. The proflavine project described in this thesis was in collaboration with him and his group members, especially Dr. Özgül Persil who I'd also like to acknowledge here.

I would also like to thank the members of crystallography labs here at Georgia Institute of Technology, both the current and the past. I would for the most like to acknowledge Dr. Tinoush Moulaei and Dr. George Lountos, both now reside at Macromolecular Crystallography Laboratory at National Cancer Institute. They have been great friends and wonderful colleagues, and I have relied on their crystallography advice for countless of times. I would also like to mention that this research would not have been possible without the contribution from other members, Dr. Derrick Watkins, Dr. Seiji Komeda, Dr. Kris Woods, Dr. Akanksha Nagpal, Dr. Linda Manning, Dr. Srividya Mohan and Dr. Chiaolong Hsiao.

I also wish to thank Professor Nicole L. LeBlanc at university of Maryland, for the use of robotic instrument for crystal screening, Professor George M. Sheldrick at Georg-August-Universität Göttingen, the author of the program *SHELX97* for helpful discussions, and Dr. Stephen Quirk at Kimberly-Clarke for dGTPase project that I worked on during my early graduate research.

My life as a graduate student would not have been completed without the opportunities in the teaching field, given by General Chemistry Teaching Faculties; Dr. Bill Baron, Dr. Toby Block, Dr. George McKelvy and my second mother, Dr. Leigh Bottomley. I believe the experience I had with you all would certainly affect my future, as it takes its shape a little by little. I would also like to express my personal appreciation to the Bottomley family, including Professor Lawrence Bottomley, one of my committee members, abovementioned Dr. Leigh Bottomley, and their wonderful daughters Christine and Marie.

Lastly, I'd like to acknowledge my parents, Shoichi and Michiyo Maehigashi for tolerating such a disobedient son. I honestly wish that completing this step of my life finally would be a good turning point for all of us.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

SUMMARY

This thesis represents results of experiments involved in ultra-high resolution x-ray crystal structures of DNA. Detailed analysis of DNA coordinates of increased accuracy, refinement of individual anisotropic *B*-factors, visualization of disorder, counterions and hydration are presumably facilitated by the increased information of ultra-high resolution structures. However, recent survey of available DNA crystal structures with the data that extends beyond 1Å has brought a few areas of considerations; local conformational flexibility of DNA and counterions and how they have been interpreted in previous studies. While the results given here demonstrate both structural and surrounding details of DNA that may be undetected at lower resolution, some of the technical aspects of interpreting ultra-high resolution data is also addressed.

First part describes the 0.96Å crystal structure of one turn of G/C rich B-DNA. Providing much improved data, thorough analysis revealed details of B-DNA conformation, as well as greater information on the local variations that may occur within such helices. Many phosphate groups are shown to assume in alternate states, and such static disorders are extended to the base pairs at the center of the duplex. One phosphate group is shown to transition from B-helical subtype B_I to B_{II}. Ultrahigh resolution structure also enabled comprehensive analysis of localized divalent cations. Many are observed to be disordered and/or partially occupied. A strong correlation observed between these counterions and static disorders within helices is also described.

Locating monovalent cations is also examined in order to address fundamental questions about the interaction of B-DNA and monovalent cations, as well as to examine the issue of charge disparity between DNA and counterions through the large majority of

published DNA crystal structures. The locations of counterions around B-DNA were determined through 0.98Å and 0.87Å resolution x-ray diffraction of crystals. These are the first examples of <1Å resolution B-DNA crystal structures within which the monovalent cations were analyzed. The majority of monovalent cations are localized in the G-tract major groove; however a few are also identified in the minor groove at TpG steps. The mode of coordination of these monovalent cations to the base functional groups and backbone of B-DNA are in essence similar, yet are observed to differ from previously determined monovalent cation sites in different sequences of B-DNA. The differences observed in coordination suggest the sequence dependence of cation localization.

Lastly, a 1.1Å resolution x-ray crystal structure of an intercalative complex between $[d(CGATCG)]_2$ and proflavine is described. Two proflavine molecules are observed to intercalate at CpG steps at the both terminals of the DNA duplex. Proflavine orients itself so that its terminal amines are toward the major groove. Proflavine molecules are in close van der Waals contacts with the adjacent C/G and G/C base pairs, positioning themselves in the pockets created by these unwound steps. The mode of intercalation by proflavine molecules is observed to be largely through stacking interaction, however several water and ion mediated contacts to the neighboring bases and phosphodiester backbone are also observed. Helical parameters of the current structure reveal that intercalation causes DNA to unwind and lengthen dramatically at the intercalation sites. These changes in helical parameters are found to be similar to the published intercalative complex of DNA with another high planarity intercalator, ellipticine.

CHAPTER 1

INTRODUCTION

Ultra-high resolution structures of B-DNA

Detailed analysis of DNA coordinates of increased accuracy, refinement of individual anisotropic B-factors, visualization of disorder, counterions and hydration are facilitated by the increased information of ultra-high resolution structures (<1Å). Despite the recent advancement in the field of x-ray crystallography, namely synchrotron radiations of increasing power, cryo-crystallography and improved detectors, the availability of ultra-high resolution DNA crystal data is still limited. Currently, there are seventeen available DNA structures (excluding mono-, di-, trinucleotides) with the data that provides resolution better than 1Å. Among these, many are in the form of more exotic DNA structure, such as Z-DNA (Wang, Quigley et al. 1979; Bennett, Krah et al. 2000; Dauter and Adamiak 2001; Tereshko, Wilds et al. 2001) and DNA Quadruplex (Phillips, Dauter et al. 1997). Two structures are of the A-form DNA with the sequence of GCGTATACGC (Egli, Tereshko et al. 1998) and CTTTTAAAAG (Han, Langs et al. To be published). Nine structures are in the B-form, with two have been published at the initial refinement stage with no solvent model assigned (Hays, Teegarden et al. 2005). Refined ultra-high resolution structures of B-DNA with solvent models include 0.83Å structure of $[d(CGCGAATTCGCG)]_2$ containing modified thymine (Egli, Tereshko et al. 1998), 0.99Å structures of $[d(CCAACGTTGG)]_2$ and $[d(CCAGCGCTGG)]_2$ (Chiu and Dickerson 2000), 0.74Å structure of $[d(CCAGTACTGG)]_2$ (Kielkopf, Ding et al. 2000) and 0.89Å structure of $[d(GCGAATTCG)]_2$ (Soler-Lopez, Malinina et al. 2000). Table

1.1 summarizes some of the structural and statistical features of these ultra-high resolution crystal structures of B-DNA.

Table 1.1:Survey of current B-DNA structures (<1.0Å)

ID[a]	Sequence	Res.(Å)	Rvalue	Alt.$_P$[b]	Alt.$_{Base}$[c]	B_{High}[e]	B_{Avg}[f]	Cation	Δq[g]
1D8G	CCAGTACTGG	0.74	0.105	12	2	19.35	6.52	Ca^{2+}	-7.8
1ENN	GCGAATTCG	0.89	0.135	4	0	41.40	9.80	Mg^{2+}	-3.5
1EN3	CCAACGTTGG	0.99	0.141	8	0	23.95	9.09	Mg^{2+}	-5.4
1EN8	CCAACGTTGG	0.99	0.118	10	2	19.53	6.35	Ca^{2+}	-9.0
1EN9	CCAGCGCTGG	0.99	0.140	6	0	22.00	8.55	Mg^{2+}	-2.4
1ENE	CCAGCGCTGG	0.99	0.121	6	0	25.59	6.64	Ca^{2+}	-6.0
1DPN	CGCGAAT[h]TCGCG	0.95	0.160	6	0	37.40	17.18	Mg^{2+}	-15.0

[a]PDB ID
[b]Number of sugar-phosphate backbone with alternate states per biological unit
[c]Number of bases with alternate states per biological unit
[e]Highest isotropic thermal factor (B-value) of a DNA base atom
[f]Average isotropic thermal factor (B-value) of a DNA duplex
[g]Sum of the charges from all charged species (accounting for partial occupancies)
[h]2'-deoxy-2'-fluoro-arabino-furanosyl-thymine-5'-phosphate
1D8G (Kielkopf, Ding et al. 2000) 1ENN (Soler-Lopez, Malinina et al. 2000) 1EN3 - 1ENE (Chiu and Dickerson 2000) 1DPN (Egli, Tereshko et al. 1998)

Reliability of ultra-high resolution B-DNA crystal structures

Alternate states model

One of the features commonly seen among published ultra-high resolution B-DNA crystal structures is the presence of alternate states (static disorders). A high occurrence of static disorders is reported among the deoxyribose and phosphate groups of B-DNA (Table 1.1). Many are presumably undetected at lower resolution, which is also indicated by the frequently observed high isotropic thermal factors (*B*-value) among phosphate groups in lower resolution structures (Kielkopf, Ding et al. 2000).

Ultra-high resolution diffraction data is more likely to allow static disorders of DNA to be resolved in the electron density with higher certainty (figure 1.1). Electron density maps become clearer with increased resolution, showing separated peaks corresponding to individual atoms. The final refined model determined by a single-crystal x-ray crystallography technique is still an average of multiple steady states that occur through the repeated lattice of a crystal. The degree of disorders among these steady states varies and may or may not be justified. The disorder depends on the quality of the data obtained, and may be undetected at lower resolution. When considering the reliability of a given crystal structure, there are several useful statistical terms relevant to the discussion.

Deviation of isotropic thermal factors

Isotropic thermal factor (*B*-value) is a measure of the atomic displacement about its average position due to thermal vibration and static disorder. In crystals of macromolecules, *B*-value may rise to 30 $Å^2$ to 40$Å^2$. *B*-value (B_j)is related to the mean square amplitude of atomic vibration (u) by $B_j = 8\pi^2 \overline{u^2}$ (Glusker 1994). For instance, *B*-value of 10$Å^2$ and 40$Å^2$ correspond to atomic displacements of approximately 0.35Å and 0.7Å respectively. Hence a region that displays high *B*-values with large deviation from the rest of DNA may indicate a possible static disorder, as well as a candidate for the alternate (multiple) states model. Assignment of an alternate state model should thus lower the isotropic *B*-value of the target region, as the range of displacement exhibited by a single state model is effectively reduced by imposing a finer grid to limit the average range of thermal motion of given atoms.

Upon surveying currently available ultra-high resolution structures, some regions are observed to display higher *B*-values that are considerably deviated from the average *B*-value of a given DNA (Table 1.1). Many are found to be of phosphate groups, and some of them are assigned with alternate (multiple) states model. While these ultra-high resolution structures of DNA display significantly lower overall *B*-values than that of lower resolution structures, such deviations from the rest of molecules within the same crystals, as well as within the alternate states region, are of concern.

Unusual profile of thermal ellipsoids (anisotropic thermal factors)

Similar observations are obtained upon inspecting thermal ellipsoids of ultra-high resolution structures of *B*-DNA. Disordered phosphate groups show considerably larger thermal ellipsoids in comparison to other phosphate groups within a given structure, and often are unusually elongated in one direction. Anisotropic representation of atomic displacement is substantiated only if high resolution (typically <1.4Å) is achieved by a given crystallographic data (Wlodawer, Minor et al. 2008). All ultra-high resolution structures of *B*-DNA that are currently available have been refined anisotropically. The atomic displacement is refined against the axes of an ellipsoid. Anisotropic refinement therefore introduces six additional parameters; three to define the length of three reciprocally perpendicular axes that relate the displacements, and three to define the orientation of these ellipsoidal axes relative to the crystal axes. Visual representations of the atomic displacement in various directions are obtained from thermal ellipsoid plots (Johnson 1976; Farrugia 1997).

Disorder in a crystal structure is frequently revealed by the shapes of the thermal ellipsoids. High anisotropic alternative model that displays substantially higher

anisotropy from the rest of the DNA indicates possible error in the refinement that may be imposed by stereochemical restraints applied inadequately. By analyzing their shapes and sizes, highly anisotropic group may be justified for assignment of alternate models. However, anisotropic temperature parameters that define thermal ellipsoids are highly susceptible to errors in the data set and refinement process. High anisotropy of an atom or group should thus make physical sense and explain its motion and the refinement process that is applied on.

The number of alternate models assigned is related inversely to the degree of atomic displacement, as alternate models impose finer grid to effectively limit the range of average atomic displacement. Highly deviated isotropic *B*-values and abnormally shaped thermal ellipsoids are often indications of misplaced atoms. Unlike small molecule crystallography, crystallography of macromolecules require a subset of geometric restraints to be applied during refinement (Hendrickson 1985; Kleywegt and Jones 1995), such as bond lengths/angles and planarity of each base present in DNA. This is because the number of observations (reflections) may simply be insufficient to refine tremendous amount of parameters involved in the complex crystallographic model. Under the control of geometric restraints, degree of freedom given to alternate model over the original model is significantly affected by the range within which alternate model is built that is defined by a crystallographer. Defining such range requires a careful analysis, because static disorder of a given phosphate group may be the result of obscure disorders accumulated over the neighboring segments of the model, such as adjacent sugars and may extend to bases. Thermal ellipsoids around adjacent segments may indicate correlated disorder as well as directionality within. The direction of atomic

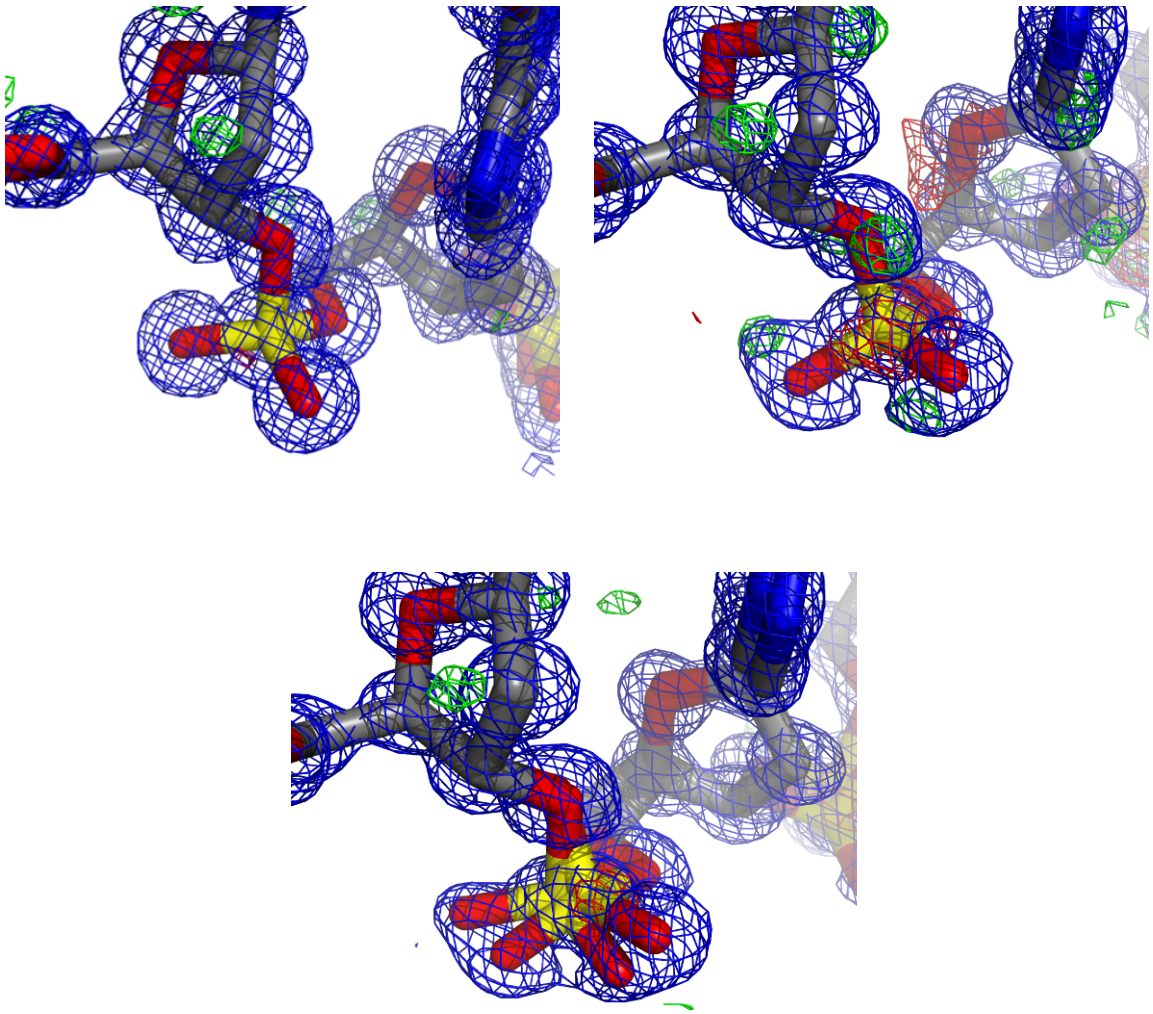

Figure 1.1: Assignment of alternate states model at ultra-high resolution. All figures are from 0.87Å B-DNA crystal structure of [d(CCAGGCCTGG)]$_2$ described in chapter 3. Sum ($2F_o$-F_c) electron density map is shown in net, colored blue and contoured at 1σ. Difference (F_o-F_c) electron density map is shown in net, colored green/red and contoured at 3.5/-3.5σ. (Top Left) A phosphate group without static disorder. Notice the defined and well separated electron densities surrounding each atom. (Top Right) Phosphate group region indicating static disorder before the second (alternative) model is assigned. Notice the elongated appearance surrounding phosphate oxygen atoms, indicative of the directions of disorder. Peaks of negative difference density, shown in green indicate the region in need of model assignment. Positive difference density shown in red appears at phosphorous, indicating excess assignment of the model. (Bottom Center) The same phosphate group after the assignment of the second alternate model and a round of refinement is shown. Notice the two alternating models are confined within the $2F_o$-F_c peaks. Difference electron densities are greatly reduced. Unless otherwise noted, figures in this chapter are rendered with Pymol v1.1 (DeLano).

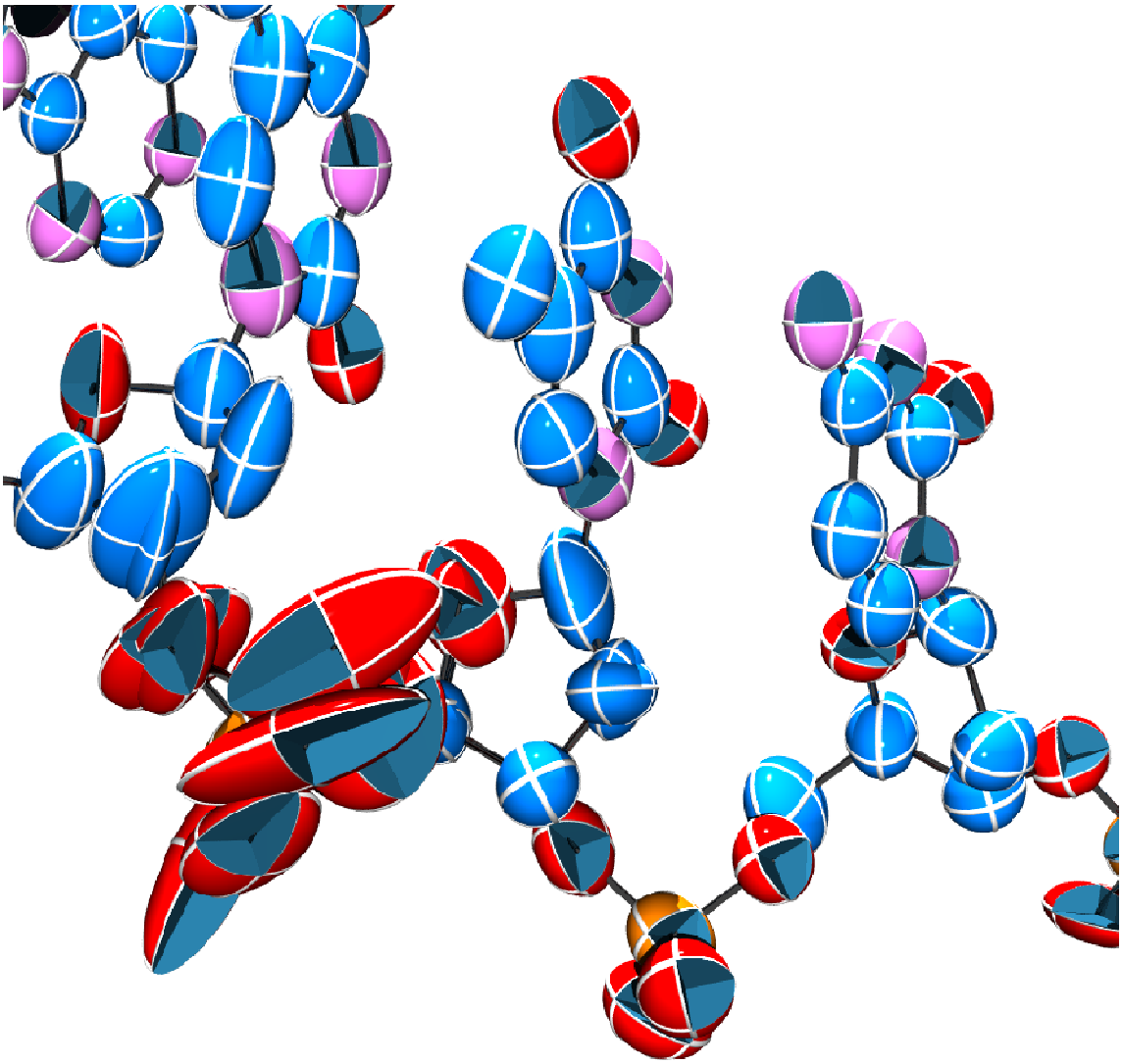

Figure 1.2: Example of an unusual correlation between alternate model assignment and thermal ellipsoids. Thermal ellipsoids of 0.89Å crystal structure of [d(GCGAATTCG)]2 (PDB ID: 1ENN) at 50% probability. T(6)T(7)C(8) steps are shown with disordered phosphate at the TpT step, which also corresponds to the highest isotropic B-value among DNA model. Notice the difference in both magnitude and shape of the thermal ellipsoids between disordered and non-disordered phosphate groups. Unusual profile of thermal ellipsoids may indicate error in model assignment, possibly due to inadequate range specified for alternate model to be built, as described in the text. Also, notice the atomic displacement and direction of disorder displayed by adjacent bases and sugars (compare to the non-anisotropic cytosine on the right). The anisotropy displayed by T(6), T(7) may indicate that the these segments are also in need of alternate model. The plots were calculated with Ortep-3 for windows (Farrugia, 1997) and rendered with Povray 3.6.

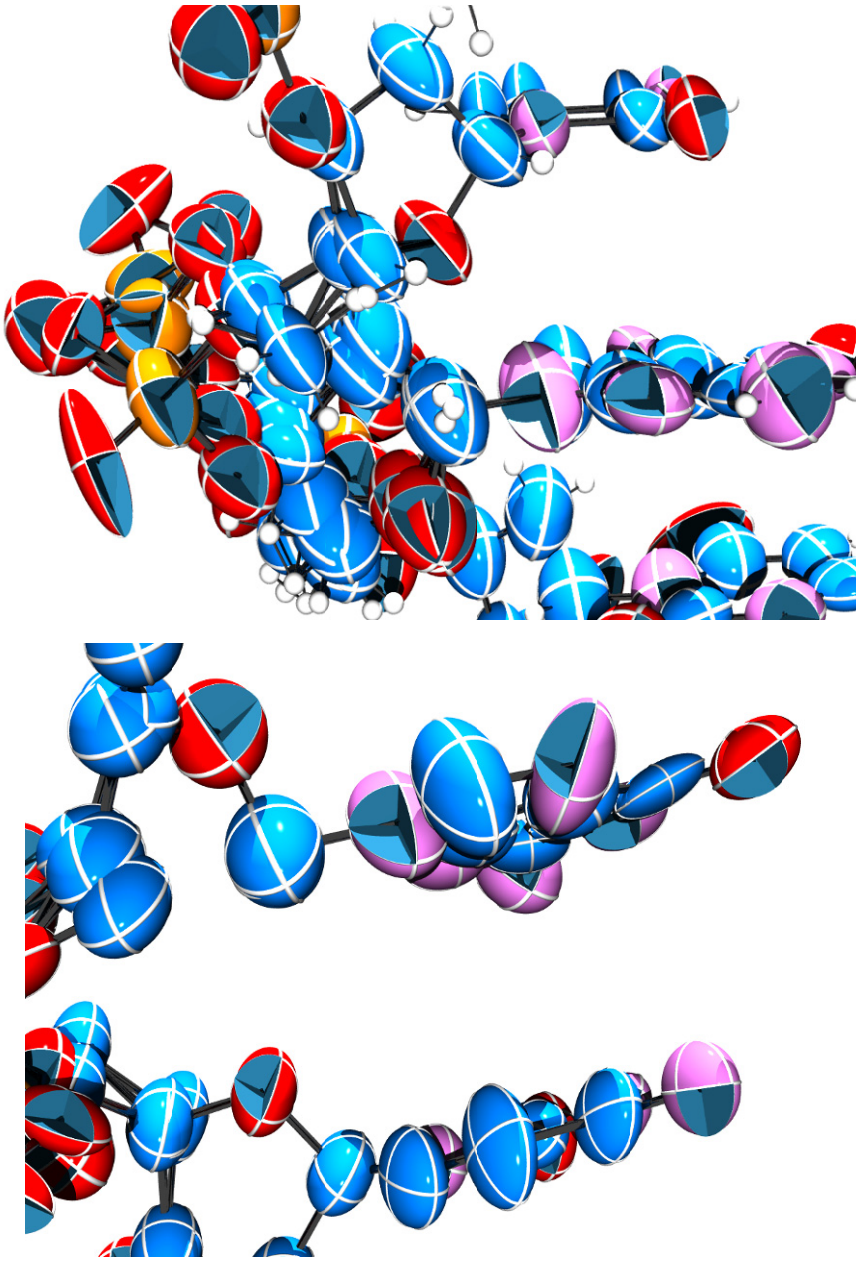

Figure 1.3: Example of an unusual correlation between alternate model assignment and thermal ellipsoids. Thermal ellipsoids of 0.99Å crystal structure of [d(CCAGCGCTGG)]$_2$ (PDB ID: 1EN9) at 50% probability. Some atoms are omitted for clarity. (Top) Three alternate models of phosphate group at the G(6)pC(7) step is shown. Notice the unusually elongated profile of one thermal ellipsoid of phosphate oxygen shown at left bottom corner. Such disparity of thermal ellipsoid may indicate that the atom is out of density, as described in the text. (Left) Adjacent bases of the phosphate group shown on left. The residue G(6) is shown at the top and C(7) is shown at the bottom, viewing into the major groove (flipped from left figure). Notice the different profile of thermal displacement observed between two bases. Mixed directionality displayed by the thermal ellipsoids of the base G(6) may indicate the base is also disordered, possibly altering propeller twist angles. However, alternate model was not assigned for G(6), therefore geometric restrains may have adversely affected on one of the alternating models, causing one phosphate oxygen to be out of density after restrained refinement. See text for more explanation. The plots were calculated with Ortep-3 for windows (Farrugia, 1997) and rendered with Povray 3.6.

displacement may be in one direction, but mixed directionality may indicate the propeller twist and/or the buckle of a given base. Inadequately defined region for alternate model may thus not produce a true representation of the static disorder, and possibly result in refinement of some atoms to be out of electron densities due to the geometric restraints.

Real-space R-factor and Real-space correlation coefficient

Real-space R-factor (RSR) and real-space correlation coefficients (RSCC) were first introduced by Jones et al (Jones, Zou et al. 1991). While conventional *R*-factor measures the global discrepancy between the experimentally obtained and calculated structure factors, RSR and RSCC are calculated to reflect the correlation between the experimental electron density map and the one generated solely from the model. Both maps are generated using the same, model-derived phase information, thus RSR and RSCC are effectively the measures of errors in map interpretations. RSR is calculated as $\mathrm{RSR}=\sum|\rho_{obs}-\rho_{calc}|/\sum|\rho_{obs}+\rho_{calc}|$, where ρ is an electron density values at grid points that cover the given residue. RSCC is calculated similarly as RSR, except the linear correlation coefficient between the two density arrays is calculated. RSCC is reported in the values between -1.0 (perfect anti-correlation) and +1.0 (perfect correlation). Both RSR and RSCC are calculated for residue by residue, therefore sudden deviation of the reported value from the rest typically indicates the trouble spot within the assigned model.

Both values can be determined using the program *mapman* (Kleywegt and Jones 1996), a part of *rave* package from Uppsala Software Factory. For a published structure, RSR and RSCC can be also obtained through external link available from Protein Data

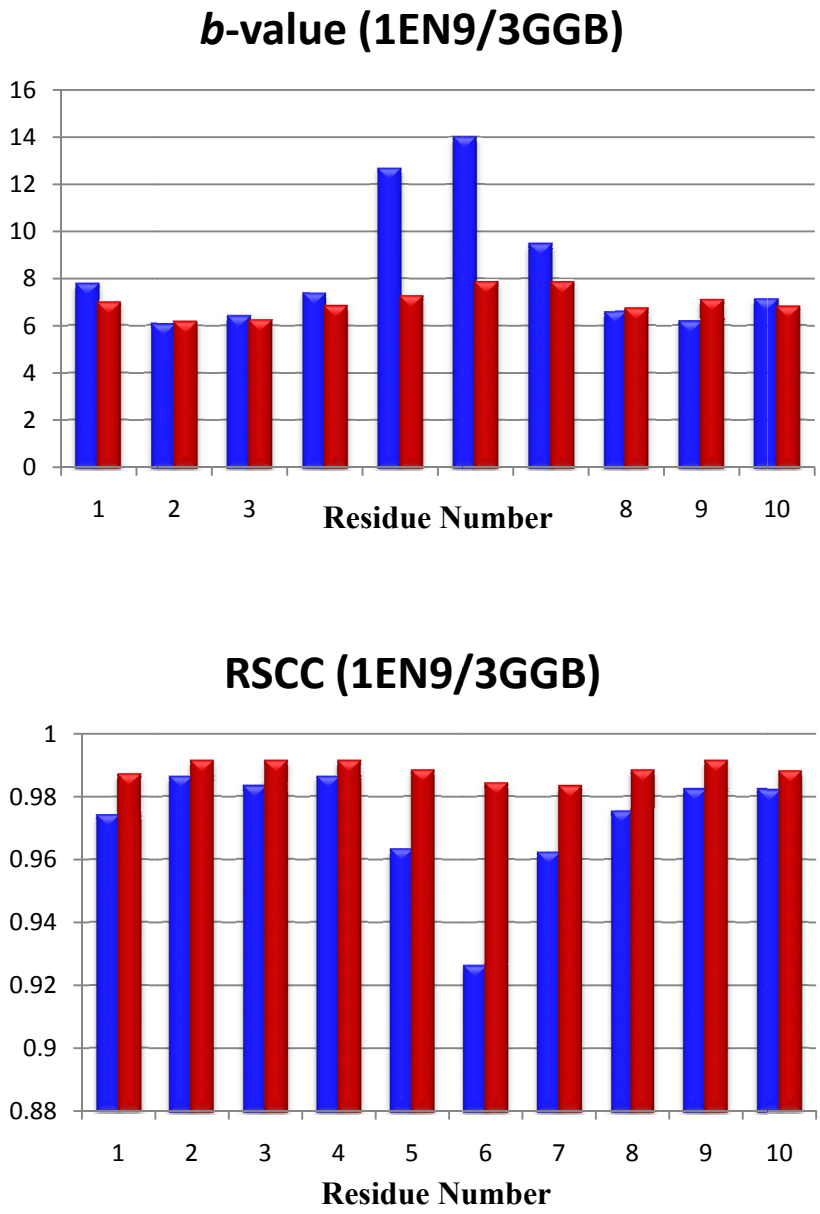

*Figure 1.4: Isotropic b-value and Real-space correlation coefficient (RSCC) plots. Each value is shown as a function of residue number. Valued of Na^+ form of [d(CCA**GGCC**TGG)]$_2$ (PDB ID 3GGB, chapter2) is shown in red. The published isomorphous 0.99Å crystal structure of [d(CCA**GCGC**TGG)]$_2$ [PDB ID 1EN9 (Chiu and Dickerson 2000)] is shown in blue. Notice the deviation of isotropic b-value at C(5) and G(6) of 1EN9, as well as sudden drop of correlation coefficient at residue G(6) of 1EN9, indicating a serious error in map interpretation. For thermal ellipsoid of the same residue, see figure 1.3. In 3GGB, the same residue [C(6) in 3GGB] is built with 2 alternate states model (see chapter 2). Isotropic b-value and RSCC for 1EN9 are obtained from Electron Density Server (Kleywegt, Harris et al. 2004), which uses the program mapman (Kleywegt and Jones 1996). Since 3GGB is at pre-released stage at the time of writing, values for 3GGB was determined in-house using the same program.*

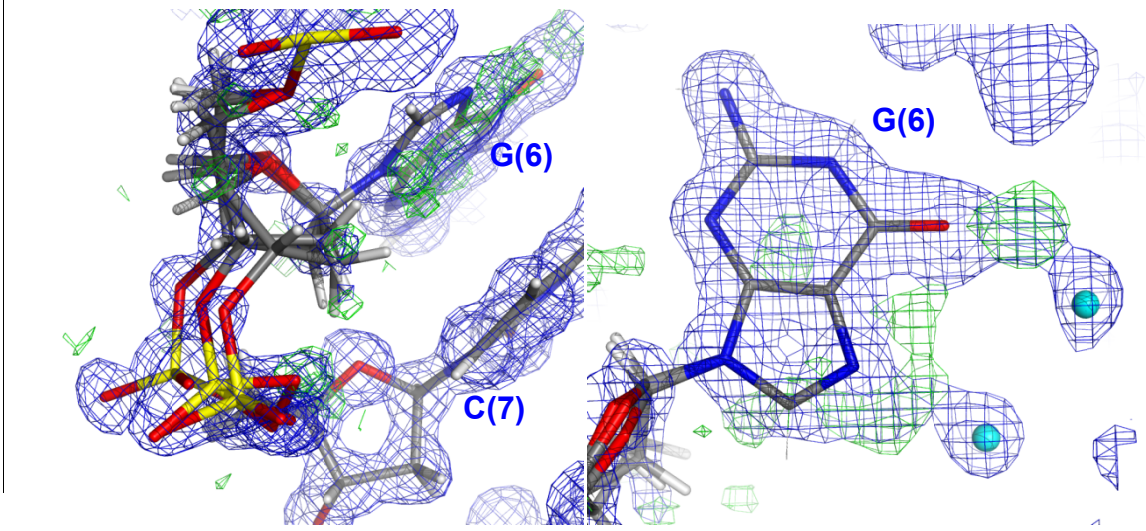

Figure 1.5: Electron density maps and final refined model of 1EN9, 0.99Å crystal structure of $[d(CCAGCGCTGG)]_2$ *(Chiu and Dickerson 2000). The structure is isomorphous to* $[d(CCAGGCCTGG)]_2$ *described in this thesis (chapter 2-3). Sum density (*$2F_o$-F_c*) is shown in blue net, contoured at 0.8σ, and difference density (*F_o-F_c*) is shown in green net, contoured at 2.5σ. Both maps were obtained from Electron Density Server (Kleywegt, Harris et al. 2004), where maps were calculated based on the deposited intensities. (Left) The G(6)pC(7) is shown, where the authors assigned 3 alternate models for the phosphodiester linkage. Hydrogen atoms are included in their final refined model. Notice the atoms that are out of density, such as C3' and O3' of some of the alternate models. (Right) The residue, G(6) is shown, where excess appearance of* F_o-F_c *peaks indicate the necessity for the second model.*

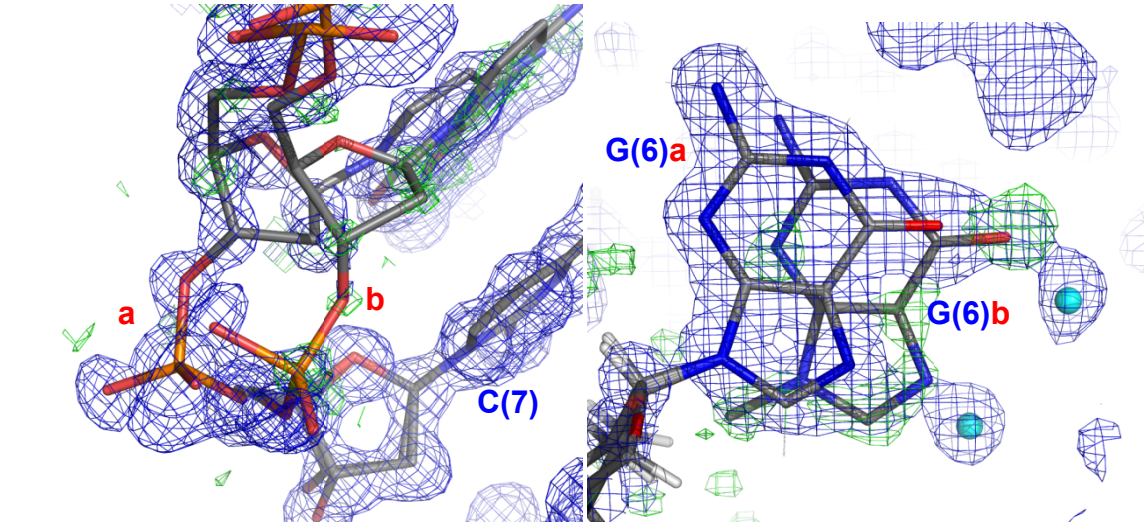

Figure 1.6: Electron density maps of 1EN9 and suggested correction to the model. The electron density maps are the same as in Figure 1.5 above. (Left) The suggested alternate model for the phosphodiester linkage at the G(6)pC(7) is shown. Model is taken from $[d(CCAGGCCTGG)]_2$ *described in later chapters from the same step, therefore cytosine residue is shown in place of G(6). Notice the second model adequately trace the* F_o-F_c *peaks appear in the region. (Right) The suggested second alternate model for the residue, G(6) is shown.*

Bank to Electron Density Server (Kleywegt, Harris et al. 2004). Figure 1.4 shows the comparison of RSCC between the Na^+ form of $[d(CCAGGCCTGG)]_2$ (3GGB, chapter 2) and a published isomorphous 0.99Å crystal structure of $[d(CCAGCGCTGG)]_2$ [1EN9, (Chiu and Dickerson 2000)] which has a different central G-tract sequence. A sudden drop of correlation coefficient at G(6) of 1EN9 clearly indicates the possible error in map interpretation during the refinement process.

Verifying a published structure through electron density maps

Figure 1.5 shows the electron density maps ($2F_o$-F_c and F_o-F_c) for 0.99Å crystal structure of $[d(CCAGCGCTGG)]_2$ (PDB ID, 1EN9). These maps were obtained from Electron Density Server that had been calculated from the deposited experimental data (structure factors, intensities). Upon inspecting the electron density maps, suspected error in map interpretation at the residue G(6) is now apparent as excess residual F_o-F_c is evident when contoured at 2.5σ. The C(7) phosphate, where the authors assigned three alternate models also show inconsistency between the models and the electron density maps, as some of the atoms are out of densities. In this case, the alternate models may be assigned in these positions in effort to reduce the excess F_o-F_c peaks that appeared adjacent to the original phosphate model. However, authors failed to interpret the adjacent residue, G(6) accurately, which in fact, strong F_o-F_c peaks surrounding the base indicate the necessity for the second alternate model of the base. As a result of inadequately defined range for an alternate model, the assigned models for the phosphate group are not truly representing the static disorder at this region. Some atoms are therefore believed to be forced out of the electron densities due to the geometric restraints.

Accounting for charge

The importance of recognizing the positions of localized cations arises, in part, from their possible role in dictating the sequence-dependent variations in DNA conformation (Rouzina and Bloomfield 1998; Williams and Maher 2000; Hud and Polak 2001). However, it is evident that the large majority of x-ray crystal DNA structures do not contain enough identified counterions to account for charge neutrality. Resolution achieved by the crystallographic data appears independent of the number of identified cations, which are needed to neutralize the negative charge of backbone phosphate (Table 1.1). Most identified cations are either Mg^{2+} or Ca^{2+}, which are easily identifiable on electron density maps through coordination geometry in moderate to high resolution crystallographic data. Identification of other cations, possibly localized to specific sites of DNA, present in crystallization solution may be difficult to achieve. DNA is crystallized in the presence of buffer, often with cacodylate acid titrated with sodium hydroxide. Na^+ is indistinguishable from water through x-ray crystallographic technique, because Na^+ has the same number of electrons as water (10 electrons) and the shape of the electron density is the same (spherical). Furthermore, the difference in ionic/molecular radii between Na^+ and water (0.95Å vs. 1.4Å) generally is not sufficient to distinguish between these two entities. Because the effective dispersion of the electrons about the atom center varies with thermal fluctuations, positional disorder and data quality, these effects obscure the difference in effective radii (Williams 2005). The low scattering Na^+ may thus not be located reliably unless by possible coordination chemistry that is often less regular and seldom are reported in *B*-DNA crystal structures (Komeda, Moulaei et al. 2006).

Locating monovalent cations in DNA crystal structures

Locating monovalent cations in x-ray DNA crystal structures can profit from the use of heavier cations with increased scattering power and their anomalous contribution to the x-ray scattering amplitudes. The advantage of cations with such properties is that they are directly observable and it is not necessary to detect subtle differences in coordination geometry. Heavier monovalent cations contribute to isomorphous differences between crystals that are grown in the presence of different ions, say Rb^+ and Na^+, or crystals soaked in a solution containing target ions for various lengths of time. This enables the detection of localized cations through isomorphous difference densities. Anomalous contribution can facilitate the detection of cations without subset of data, similar in the single anomalous dispersion (SAD) experiment. Examples of such crystallographic studies include the use of thallium (Tl^+) ion as a probe to determine monovalent sites in the tetrahymena ribozyme P4-P6 domain (Basu, Rambo et al. 1998), the bovine insulin (Badger, Kapulsky et al. 1994; Badger, Li et al. 1994), a duplex *B*-form DNA dodecamer CGCGAATTCGCG (Howerton, Sines et al. 2001) and intercalative complex of $[d(CGATCG)]_2$/adriamycin$_2$ (Howerton, Nagpal et al. 2003). When considering its application in a biological system, Tl^+ is an effective K^+ mimic. Tl^+ and K^+ have similar properties including ionic radii (K^+ = 1.33Å, Tl^+ = 1.49Å), irregular coordination geometries and enthalpy of hydration (K^+ = -77kcal/mol, Tl^+ = -78kcal/mol). NMR shows both ions bind and stabilize DNA G-quartets (Basu, Szewczak et al. 2000) and activate pyruvate kinase (Loria and Nowak 1998). A binding assay shows Tl^+ and K^+ are nearly interchangeable in sodium-potassium pumps (Pedersen, Nielsen et al. 1998). Marcus and Hosey showed that thallium is also an activator of fructose-1.6-

bisphosphatase (Marcus and Hosey 1980) and crystallographic evidence supports such finding (Villeret, Huang et al. 1995).

The use of anomalous contribution was further exploited with group I cations (K^+, Rb^+ and Cs^+) because, although they are weaker than that of Tl^+, these metals also exhibit anomalous contribution. Egli and co-workers determined the crystal structures of modified *A*-form decamer of GCGTATACGC with K^+, Rb^+ and Cs^+ (Tereshko, Wilds et al. 2001). Crystal structure of the pig renal Na^+/K^+ pump was determined with bound Rb^+ (Morth, Pedersen et al. 2007). Isomormous difference peak was used to detect Rb^+ bound site in the ribosomal subunit (Klein, Moore et al. 2004).

In later chapters, ultra-high resolution crystal structure of $[d(CCAGGCCTGG)]_2$ in three derivative forms, Na^+ (0.96Å), Tl^+ (0.98Å) and Rb^+ (0.87Å) are presented in detail. The results of the experiments demonstrate the effectiveness of the use of Rb^+ and Tl^+ to locate monovalent cation sites within the crystals.

CHAPTER 2

CRYSTAL STRUCTURE OF ONE TURN OF G/C RICH B-DNA REVISITED

Abstract

The crystal structure of one turn of G/C rich B-DNA, $[d(CCAGGCCTGG)]_2$ is revisited in this work. The improved data that extends to 0.96Å reveals greater details on local variations that occur within such helices. Seven out of nine phosphate groups have alternate states, and such static disorders extend to the entire base pairs in the central GpC step. The geometry of the sugar-phosphate backbone also alters between two alternate states. One phosphate group transitions from *B*-helical subtype B_I to B_{II} as DNA alters its state from A to B. Ultrahigh resolution structure also enables comprehensive analysis of counter-ions. All hexaaqua Mg^{2+} ions in the refined structure are modeled with multi states and/or partial occupancies. A strong correlation observed between these counter-ions and static disorders within helices is also described.

Introduction

Within over 700 DNA structure deposited to Protein Data Bank (PDB), it is without doubt that the most studied structures are the DNA fragments of Dickerson-Drew Dodecamer (DDD), $[d(CGCGAATTCGCG)]_2$ and its variants. As of the time of writing, 106 DDD structures are available out of the 731 DNA structures total, excluding ~40 closely related DDD variants that contain A-tract at their center of duplex (eg. ATAT instead of AATT). However, despite the popularity of DDD and its variants, they share

the common A/T rich centers causing a characteristically narrow minor groove. Hence, the importance of the study for more generic sequence of B-DNA has been suggested (Heinemann and Alings 1989). Generic DNA refers to the sequence that does not contain either A-tract or the sequence entirely composed of G and/or C, therefore more closely resembles the vast majority of genomic sequence. Their earlier work of $[d(CCAGGCCTGG)]_2$ crystal structure showed a straight form of B-DNA double helix, with its helical parameters more similar to that of standard B-DNA than A-tract DNA. Its biological relevance is apparent; the sequence is more G/C rich (80%) and contains a palindromic sequence of CCAGG. The sequence is long recognized as a mutational hotspot in *E. coli*. Briefly, the methylation of the two central cytosine residues occurs as part of the restriction/modification (RM) system that is a crucial part of defense against invading viral DNA. The resulting 5-methylcytosine is then the subject of hydrolytic deamination to thymine (Zell and Fritz 1987; Walsh and Xu 2006). Also, $[d(CCAGGCCTGG)]_2$ contains only isolated A/T base pairs as opposed to the A-tract of DDD, potentially providing detailed insights of a more generic form of G/C rich B-DNA.

Heinemann and Alings' structure was the highest resolution reported at the time of its publication (1.6Å) among other available DNA structures with 8 or more base-pairs. Despite of its rather high resolution data in even today's standard, the solvent regions of the structure may not have been resolved entirely. The original x-ray structure of $[d(CCAGGCCTGG)]_2$ contains only two triethylammonium ions (TEA^+) per asymmetric unit (single strand). The crystal solution contains a rather high concentration (250mM) of magnesium chloride; however none of the hydrated Mg^{2+} was identified. Hence, as the authors point out, there's clearly a charge imbalance between DNA and

identified cations (-18 vs. +4 *Coulombs*). Furthermore, a total of only 41 water molecules are identified (per ASU), which seems far less solvent molecules than expected when considering approximately 40% solvent contents of the crystal (V_m=2.04).

It is apparent that the much improved crystallographic data is in necessity, desirably that extends to ultra-high resolution (>1Å). The important role of the counter-ions/solvents in DNA conformational behavior and interactions of DNA has been long discussed (Saenger 1987). Atomic resolution of crystallographic data would enable the exploration of such feature and further provide useful insight, as well as adressing such discrepancy of charge imbalance seen in a lower resolution structure. Despite the recent advancement in the field of x-ray crystallography, namely synchrotron radiations of increasing power, cryo-crystallography and improved detectors, the availability of DNA crystal data that extends to ultra-high resolution is still limited. Currently, there are 17 available DNA structures with data that extend to 1Å and beyond. Among these, many are in the form of more exotic DNA structure (eg. Z-DNA) and less than half are of the B-form DNA. Moreover, there are only 4 variations of sequence patterns within (See Chapter 1) the available ultra-high resolution B-DNA structures.

Here, we re-visit the structure of one turn of G/C rich B-DNA, $[d(CCAGGCCTGG)]_2$. With the improved data that extends to the ultra-high resolution (0.96Å), it has enabled more comprehensive analysis of hydration, counter-ions and DNA coordinates with higher accuracy. With the same crystal lattice and similar unit cell dimensions, the DNA coordinates are essentially isomorphous with the original structure of $[d(CCAGGCCTGG)]_2$. However, what is apparent is the greater detail on local variations that occur within such helices. For example, the final refined model reveals

multiple static states of the DNA that extends to the two entire base pairs. Local conformational disorders within helices have been reported previously, however most are contained within the sugar/phosphate backbones of the DNA. To the best of knowledge, multiple alternating states that extend to the entire base pairs are yet unseen, even in the previously reported isomorphous crystal structure with higher resolution at 0.74Å (Kielkopf, Ding et al. 2000). While this work may serve to compliment the original detailed work by Heinemann and Alings, the improved data also redefines the range of structural variation that can occur locally. Atomic resolution data also provides the precedent view of the solvent region, revealing more numbers of counter-ions and water molecules and their interactions with the DNA helix.

Materials and Methods

Crystallization and Data Collection

Reverse-phase HPLC purified d[CCAGGCCTGG] (Integrated DNA Technology) was annealed slowly from 90°C in a water bath upon reconstitution to the desired concentration in nano-pure water. Crystals of $[d(CCAGGCCTGG)]_2$ were grown by hanging drop vapor diffusion method, from a crystallization drop initially containing 0.23mM (single strand base) of d(CCAGGCCTGG), 12.5mM magnesium acetate, 25mM sodium cacodylate buffer (pH6.5) and 20% (v/v) 2-Methyl-2,4-pentanediol (MPD). The crystallization drop was equilibrated against a reservoir solution containing 25mM magnesium acetate, 50mM sodium cacodylate (pH6.5) and 40% (v/v) MPD at 4°C. Plate-like crystals with dimensions approximately 0.2 x 0.1 x 0.05mm^3 appeared, typically in 2 -3 days at 4°C. Despite their relatively small size, the crystals displayed high birefringence characteristics and their edges appeared sharp.

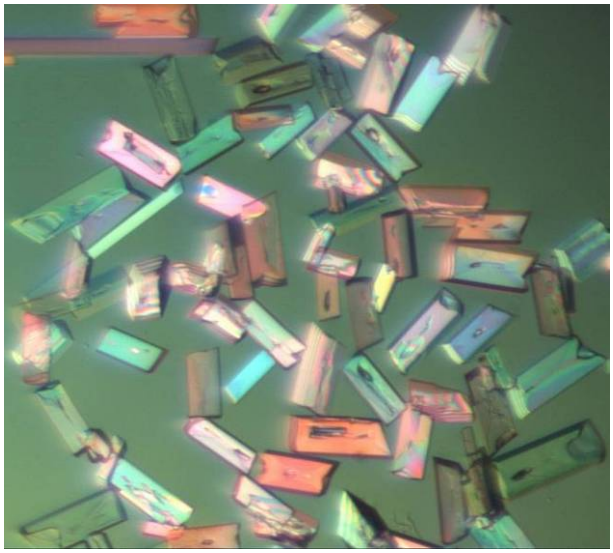

Figure 2.1: Crystals of [d(CCAGGCCTGG)]$_2$ with the presence of Mg^{2+} ions. The average dimension of crystal is approximately 0.2 × 0.1 × 0.05mm^3. Notice the sharp edges and high birifringency displayed by the crystals. The crystals are shown under polarized light.

Crystals were looped with 0.1-0.2mm 20micron nylon cryoloop (Hampton Research, California, USA), followed by flash-freezing in liquid nitrogen and kept frozen until the data collection. X-ray diffraction data used for the refinement was collected at beamline 22-ID in the SER-CAT facilities at Advanced Photon Source (Argonne, IL, USA) for a total of 360° of data with 1° oscillation using MAR 225 CCD detector (Marresearch GmbH, Germany) with 1.0Å radiation (Table #). The crystal was cryogenically cooled at -160°C during data collection. These crystals diffracted strongly, with frequently overloading the detector even at the mechanically applicable minimum exposure time (1 second) at each oscillation, thus a thin metal foil was applied to the path of the incoming beam to minimize the damage to the detector. A total of 92542 reflections through 360 frames were collected, indexed and reduced to 14269 unique reflections at the maximum resolution of 0.92. The complex of magnesium-[d(CCAGGCCTT)]$_2$ crystallizes in the monoclinic space group C2, with unit cell

parameters of a=32.14Å, b=25.17Å, c=34.09Å, α=γ=90° and β=116.25°. The table 2-1 summarizes the data collection and refinement statistics.

Refinement

The phase determination was carried out by a molecular replacement method with the DNA coordinates from d(CCAGGCCTGG)$_2$ at 1.6Å [NDB entry BDJ017 (Heinemann and Alings 1989)] using the program *CNS version 1.1* (Brunger, Adams et al. 1998) along with the parameters of Berman and co-workers (Berman, Zardecki et al. 1998). The improved starting model that consists of DNA coordinates from the program *CNS 1.1* through a several round of simulated annealing and refinement was directly fed to the program *REFMAC5* (Murshudov, Vagin et al. 1997). Care was taken in terms of preserving the flagged intensities for Free-R test between these two programs. The original intensity file (.sca) was converted to MTZ using the program sca2mtz first, and then this MTZ output was processed by the program Mtz2various to be used in *CNS*. Both utilities are available in *CCP4 program suit* (Bailey 1994).The same MTZ file was used for the refinement process by the program *REFMAC5*. Hydrogen atoms were added in the riding positions during the refinement process (with no output to the final PDB), and all atoms in the model were refined anisotropically. The refinement process was continued until the convergence of refinement statistics was reached (Table 3.1), while R-free and thermal ellipsoids were monitored to avoid over refinement. Thermal ellipsoids (figure 2.2) were computed with *ORTEP-3 for Windows* (Farrugia 1997).

Multi-state model of DNA

When building multi-state models for DNA coordinates, a single state for the entire strand was assigned and refined first (alternative state A), then the second state

(alternative state B) was assigned and refined based on the discrete appearance of F_o-F_c peaks adjacent to the original model. Finally, all atoms in both alternating states were refined anisotropically. The program *SHELXL* (Sheldrick 1997; Sheldrick 2008) was utilized in the initial assignment of occupancies in regions where alternative states were observed (see appendix A). This was followed by manual adjustment of the estimated occupancy by carefully inspecting corresponding $2F_o$-F_c and F_o-F_c peaks, followed by several round of refinements to finalize the model.

Identification of hydrated magnesium ions

Hydrated (hexaaqua) Mg^{2+} ions were added to the final model where the $2F_o$-F_c *Fourier* and the appearance of corresponding F_o-F_c *difference Fourier* electron density maps both satisfied the expected coordination geometry of the hydrated magnesium complexes. Typically, a fully or partially occupied hexaaqua Mg^{2+} displays a cluster of six surrounding peaks (water) to the magnesium center in octahedral geometry. The bond distance between each coordinated water and Mg^{2+} center is approximately 2.1Å. Each model was built manually with the program *Coot* (Emsley and Cowtan 2004), followed by rounds of refinement.

Hydrated Mg^{2+} ions with relatively high occupancies (>50%) were added to the model first, as they display clear peaks showing the expected coordination geometry. Assignment of the less occupied hydrated Mg^{2+}, especially at the disordered region was halted until the model had improved in the later stages of the refinement. In some extreme cases, where the occupancy of the Mg^{2+} complex is quite low (<20%) and highly disordered, the site was initially assigned with low occupied/disordered water molecules. The improved and more defined appearance of both $2F_o$-F_c and F_o-F_c peaks in later

stages of the refinement helped identify the site to be occupied by a Mg^{2+} complex with 3 discrete alternative states. This "water-first" approach is particularly useful in identifying the region where assignment of multiple solvent molecules would create a conflict (too close contact distance). It also helps in visualizing the neighboring spots where additional modeling is required, by generating more defined F_o-F_c peaks after the modeling /refinement. In any case, all of the assigned Mg^{2+} complexes were treated as "whole" with the same occupancies for all atoms within the complex.

Despite the increasing trend of assigning partially hydrated Mg^{2+} complexes or non-uniformed occupancies within the same complex, such irregularity is not reported in the current structure. Even at atomic resolution provided, such phenomena could not be reasonably explained by possible static disorder, for instance, a rotation around the axis that runs through the complex. The program *SHELXL (SHELX-97)* was utilized in initial assignment of occupancies to the regions where multiple states were observed (see Appendix A). The occupancy for each complex within the same disordered site was carefully adjusted accordingly. As a result, the sum of occupancies among multiple magnesium complexes within the same disordered site is not necessarily 100%. This is due to the existence of additional disordered water molecules within the same site, which takes an additional alternate state or as a result of occupancy adjustment because of the appearance of positive difference density (Fo-Fc <0) that clearly indicates the excess assignment of the model.

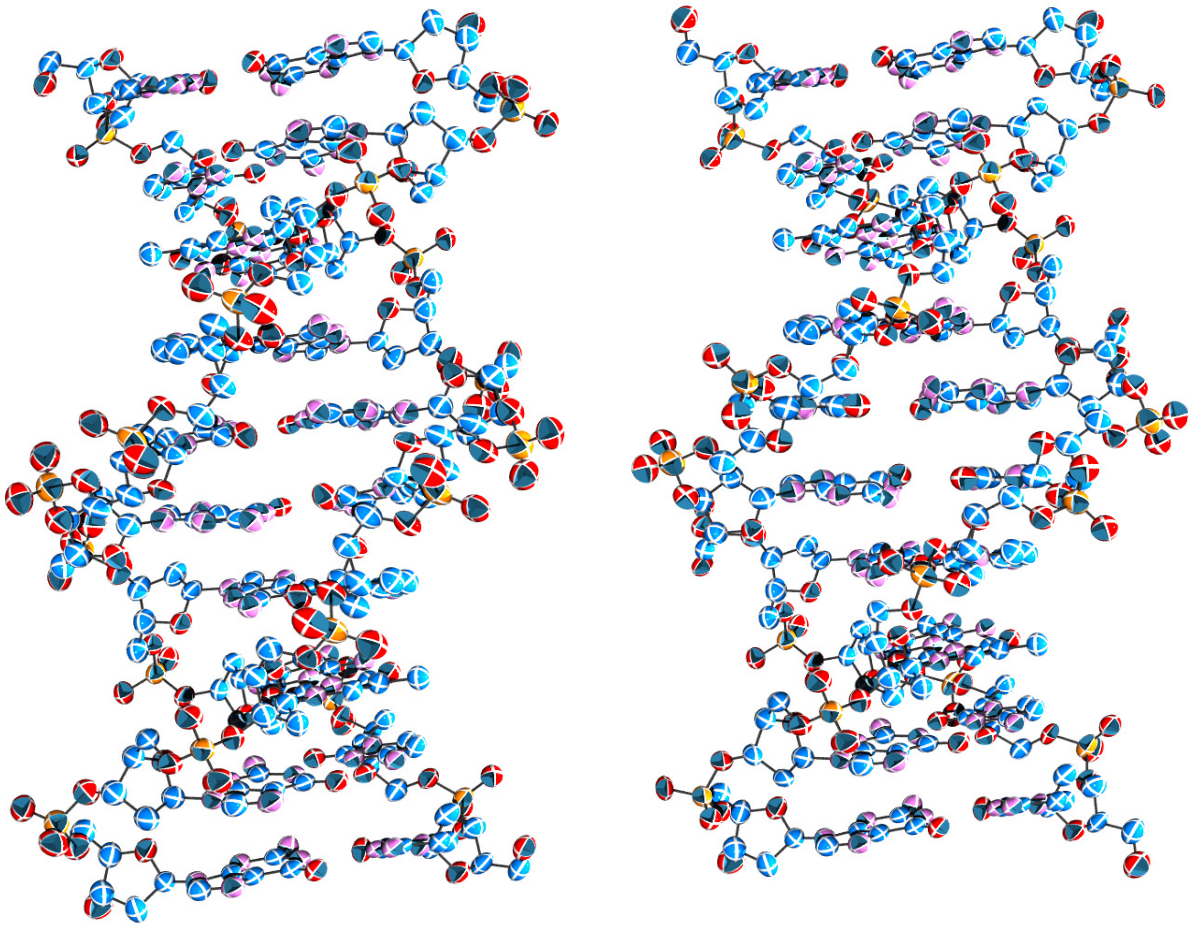

Figure 2.2: Thermal ellipsoids of [d(CCAGGCCTGG)]$_2$ at 50% probability. (Left) Alternative state "A" (Right) alternative state "B". Both viewing into the minor groove. In general, sugar/phosphate backbones show larger anisotropy indicating their thermal motion. However, the current structure does not display great difference in anisotropy between disordered regions and non-disordered region, since multi states model are built to compensate such issue. The plots were calculated with Ortep-3 for windows (Farrugia, 1997) and rendered with Povray 3.6.

Table 2.1: Crystallographic and Refinement Statistics

Data Collection	
Source wavelength (Å)	0.99997
Space group	C2
Unit cell	a = 32.136Å b = 25.172Å c = 34.094Å
	α=γ=90° β=116.25°
Resolution range (Å)	16.0-0.96
Number of reflections (all)	92542
Number of unique reflections	13935
Completeness (%)[a]	92.44 (50.27)
Average I/σ (I)	53.4
R_{merge} (%)[b]	7.9
Refinement statistics	
DNA (asymmetric unit)	d(CCAGGCCTGG)
Number of DNA atoms	287
Number of water molecules	82
Number of hexaaqua Mg ions	5
R_{work} (%)[d]	11.8
R_{free} (%)[e]	13.6 (1087 reflections)
RMS deviation of bonds from ideal (Å)	0.010
RMS deviation of angles from ideal (°)	1.880
Average isotropic B value	8.610
PDB ID code	3GGB

[a] *The values in parentheses refers to the highest resolution shells*
[b] $R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma I$, *where I = observed intensities and* $\langle I \rangle$ *= mean intensity obtained from multiple observations of symmetry-related reflections and rejections.*
[c] *Includes partially occupied species as well as molecules on special position. Disordered species with the same residue number are counted as single molecules.*
[d] $R_{work} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$, *where Fo and Fc are the observed and calculated structure factors, respectively. Reflections flagged for Free R test (7.8%) are excluded from the calculation. R factor for all reflections is determined to be 11.97%.*
[e] R_{free} *= defined by Brünger (Brunger 1992)*

Results and Discussions

A total of 13935 reflections in the resolution range 16.1-0.96 were used in the refinement. There is no exclusion of the reflections by sigma-cut, only by the limitation of the maximum resolution used due to the low completeness of the resolution bin at higher resolution. The final *R factor* is determined to be 11.97% for all data. The asymmetric unit contains a single strand of d(CCAGGCCTGG). As customary, bases are number 1-10 in the first strand, 11-20 in the other throughout the text. Two strands are related by the crystallographic 2-fold axis of C2 space group to complete a biological unit. All atoms for DNA have been modeled in and coordinates are well defined, as indicated by the clean and defined electron density maps (Figure 2.3) and the low isotropic thermal factors determined (<10 for DNA coordinates). The final model includes frequent disordered regions mostly in its sugar/phosphate backbone region, but also extends to the base pairs at the center of the duplex. Additionally, close inspection of both $2F_o$-F_c and F_o-F_c electron densities revealed five partially occupied hydrated (hexaaqua) Mg^{2+} ions, and a total of 93 water molecules (including partially occupied) per asymmetric unit.

DNA Conformation

Multiple states of DNA

High resolution and the quality of the current data set allowed accurate modeling of DNA molecules, including the discrete alternative states of DNA as shown in figure 2.4. The existence of multiple states taken by $[d(CCAGGCCTGG)]_2$ is evident in the current structure through the appearance of excess F_o-F_c peaks contoured at 3.5σ adjacent to the sugar/phosphate backbones. The final refinement model reveals a total of seven out

of nine phosphate groups per asymmetric unit (single strand) having at least two distinct alternative states. In most cases, multiple states also extend to the adjacent sugar moieties. The occupancies estimated for each state range from 25% to 75%. As described earlier, the occupancy for each state was initially estimated by the program *SHELXL*, followed by manual adjustment upon inspection of the surrounding electron densities.

In spite of the high occurrence of disorder observed in the current structure, isotropic thermal factors (*b*-value) of almost all atoms within the DNA fall into the low value range (<10). This observation suggests the existence of static disorders (in contrast to random disorders), which is described by the multiple discrete states occurring within the crystal lattices. The lack of significant difference in isotropic thermal factors between disordered and non-disordered regions indicates the accuracy of the multi-state models. Occupancies and isotropic *b*-values for each base and sugar/phosphate backbone are summarized in figure 2.5. In most cases, there appear to be one preferred state that ranges between 60% to 75% that is coupled with a second state, presumably less preferred with lower occupancy. The only exception to this is the alternate states observed at G(9)pG(10) step (A:B=50:50%). The high occurrence of multiple states also extends to the hexaaqua magnesium (Mg^{2+}) complexes, which is explained in the later section.

Global helix structure

The final refined model of $[d(CCAGGCCTGG)]_2$ shows its helical axis virtually straight, and the difference between alternate states "A" and "B" is negligible (Figure 2.16). The overall axis bend angle measured between vectors composed of the first and last helix segments is 0.12°/2.14° (alternate state A/B). The difference between the path

length of the curved helical axis and its end-to-end (straight axis connecting two ends) is 0.21Å (0.66%)/0.10Å (0.34%) for alternate state A/B.

On the other hand, the difference between the two alternate states is more pronounced in groove width. The program *CURVES 5.3* displayed the widening of the minor groove in the G-tract, as the result of the alternate conformation, with any of the cases where phosphorous, C1' or C4'atoms defining backbone splines between the measured minor groove width. Groove widths calculated by the program *CURVES* are plotted in the figure 2.6. The plots show the trend where the widening of the groove is actually toward the floor of the groove, as moving toward the center of the duplex. Through the measurements of the groove width using phosphorous atoms, the minor groove is the most widened at residues G(4) and C(7). This is followed by the widening of the mid-depth of the groove at G(4)pG(5) and C(6)pC(7) steps, evidenced by the measurement with C4' atoms. From the measurement with C1' atoms, the floor of the groove is most widened at the center of the duplex, at G(5)pC(6) step.

Extreme structural variations in the central G-tract

The most extreme structural variations are observed at the center of duplex with the sequence of GGCC (residue 4-7). The furanose ring of C(6) and the following phosphate group at C(6)pC(7) step greatly shift toward the major groove, with the measured distance of >2.0Å between the phosphorous atoms and C3' between its alternative states A and B. The great extent of disorder was evident because of the cluster of strong F_o-F_c peaks (>5σ for phosphorous and O5'atoms in state B, >3,5σ for OP1 in state B; all state B positions determined later) that appeared in proximity to where the original model was built.

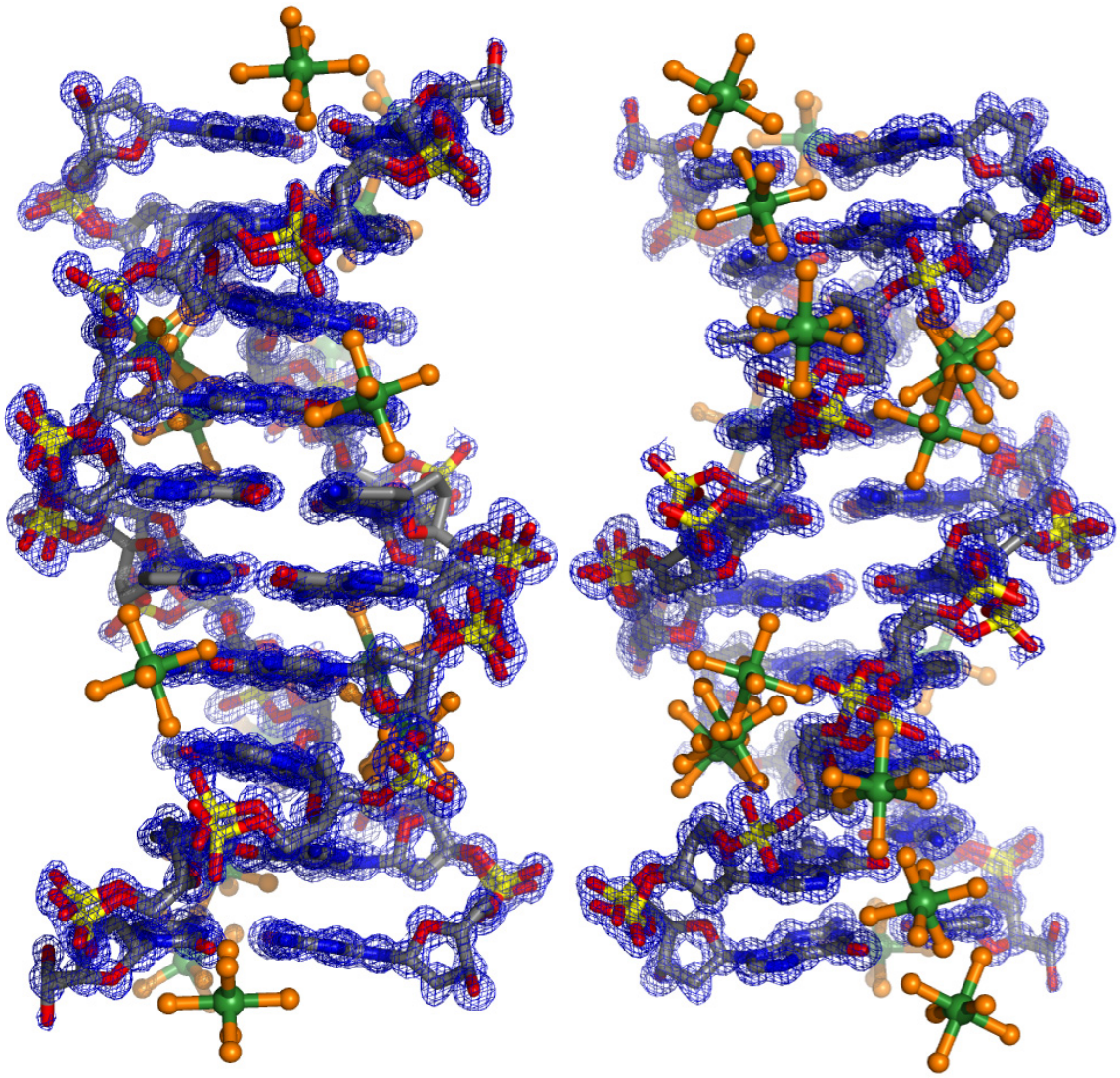

Figure 2.3: Overall view of the Magnesium-[d(CCAGGCCTGG)]$_2$ complex. Viewing into the major groove (left) and the minor groove (right). DNA strands are shown in stick models, colored with the CPK coloring scheme. Sum electron density ($2F_o$-F_c shown in blue net) is contoured at 1.0σ around DNA only for clarity. Hexaaqua magnesium complexes are shown in ball and stick models, with Mg^{2+} colored green and coordinating water molecules colored orange. Unless otherwise noted, all figures in this chapter were rendered with Pymol v1.1 (DeLano).

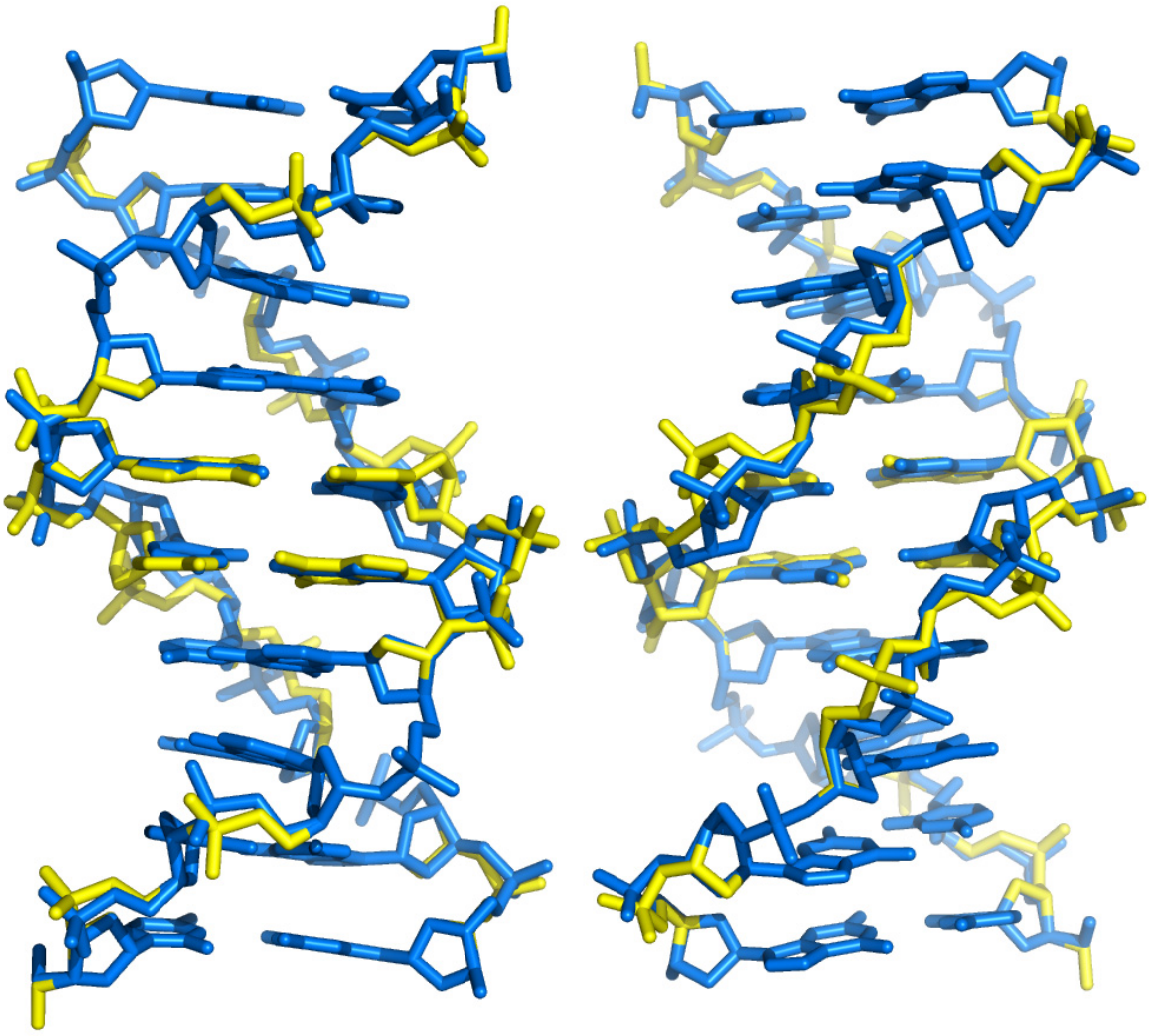

Figure 2.4: Global view of alternate states of [d(CCAGGCCTGG)]$_2$. Viewing into major groove (left) and minor groove (right). DNA strands are shown in stick models. Alternative state "A" as well as where a single state is observed are colored blue, and alternative state "B" is colored yellow. Notice the change in minor-groove width at the center of the duplex upon state transition.

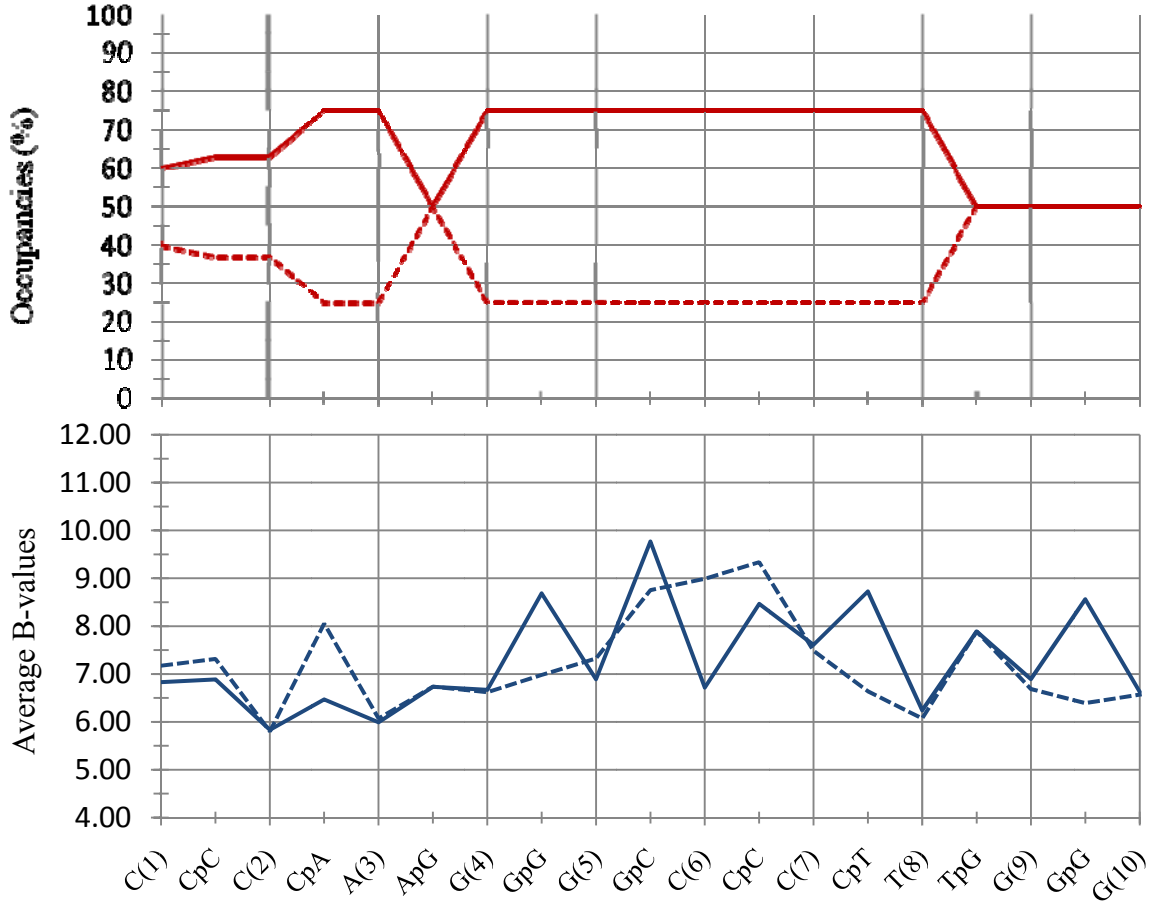

Figure 2.5: Plots of average isotropic B-values and occupancies. The solid line represents values for alternative state "A", while the dotted line represents those for state "B". Values are displayed for deoxynucleosides (eg. C(1)) and phosphate groups (eg. CpC) separately, as indicated on the common x-axis for both plots at the bottom. Both connecting O3' and O5' are considered as part of the connecting phosphate group except for the terminal oxygens (terminal O5' and O3') at both ends of the strand, which are combined to the terminal nucleaosides. Note that A(3)pG(4) and T(8)pG(9) do not include any atom with alternate state, hence occupancies are 1:1 with the same average thermal factor for both states.

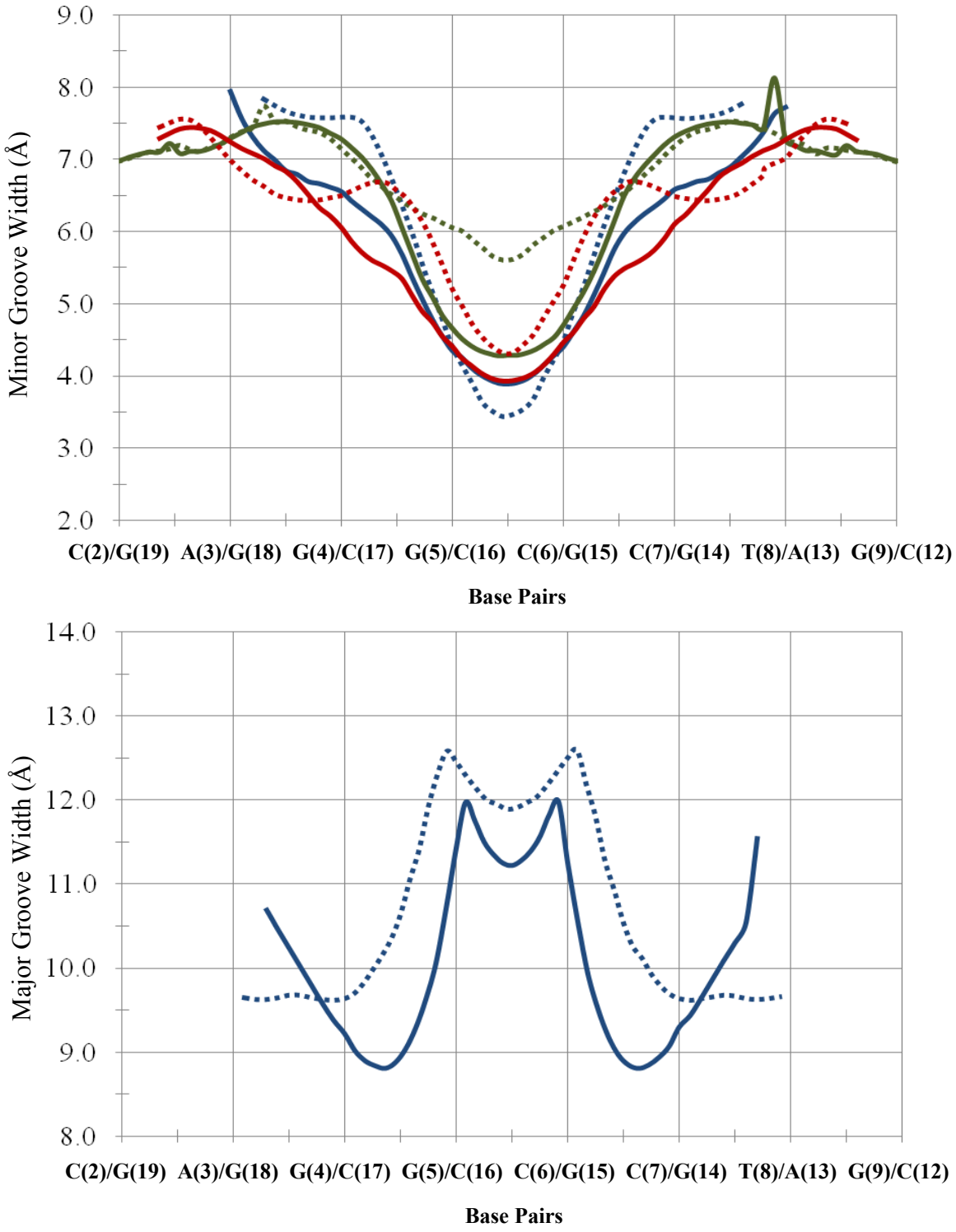

Figure 2.6: The minor and major groove widths calculated by the program CURVES 5.3. Groove widths are shown on the Y-axis for the base pairs in the duplex, strand A/B is displayed on X-axis. The solid lines represent alternative state A, while dotted line represents state B. The atoms defining the backbone splines between where groove widths were measured are phosphorous (blue), C1' (green) and C4' atoms (red).

Corresponding $2F_o$-F_c peaks also overlap with these F_o-F_c peaks indicating that the assignment of an additional model was necessary. These cluster of peaks are obviously too close (<1.5-2.3Å) to be solvent molecules, such as water molecules that frequently appear around phosphate groups (Schneider, Patel et al. 1998). Figure 2.7 and 2.8 show a close up of the phosphate group at C(6)pC(7) step with sum densities after the refinement overlapped with the difference densities at pre-refined stage (before the alternate model was built in).

The strong positive difference densities also appeared around the original model (>5σ around original P), indicating the excess assignment of the original model (100% occupancy). However, it was realized that the assignment of the alternative state cannot be done reasonably by a simple-shift/slide of the original phosphate group as with the other disordered phosphate groups in the duplex. For a given phosphate, such approach would have caused an unrealistic geometry with the adjacent sugars. Also, despite the elongated and definite $2F_o$-F_c peaks displayed at connecting O5' position (>2σ) suggesting the slide/shift of the phosphate group, the original O3' position did not display such elongated sum densities. This observation suggests that the static disorder is not only contained only within the phosphate group at this region. Therefore, an approach to build an alternative state for entire sugar moiety and base of C(6) became appropriate.

Such a rather extreme approach was further supported by the following observations. (i) There are two strong F_o-F_c peaks (>3.5σ) and weak, but distinctive $2F_o$-F_c peaks near C2' of C(6) with the distance of 1.2-1.4Å pointing toward the major groove (Figure 2.9). These peaks are too far away to be originated from riding hydrogen atoms at this position, and despite the resolution limit to which the current data set extends, such

strong peaks are not expected for hydrogen atoms (Kielkopf, Ding et al. 2000; Wlodawer, Minor et al. 2008). These peaks were later assigned as O3' and C3' of C(6) in state B. (ii) The persistent F_o-F_c peaks (>4σ) also appeared around the C(6) base toward the major groove, also suggesting the shift of the base to the same direction (Figure 2.10).

The extensive shift of the residue C(6) and the following phosphate group at C(6)pC(7) step is extended to the residue G(5), where an alternative state was also built for the entire residue. This is expected because its crystallographically unique G(15) in the opposing strand forms Watson-Crick base pairs with C(6) and to conserve the hydrogen bonding in between. The estimated occupancies for the second conformational state is 25% overall, which accounts for the weak $2F_o$-F_c peaks for some of the carbon atoms (eg. C2' and C5 of cytosine 6 in state B = ~0.7σ). In general, furanose rings show the weakest electron densities due to its flexible nature and longer bond length because of the lack of double-bonds.

In the current structure, crystallographic 2-fold symmetry relates both strands, or the top half of the duplex with the bottom half. As a result, the entire bases of both GC pairs in the middle of the duplex (residues 5-6 and 15-16) were assigned with two alternate states. Within these GC pairs, the larger movement is seen in cytosine residue, while the movement of the guanine base is best described as a twisting toward the major groove while being anchored at its C1'. This is also described by the small difference in guanine sugar positions between the two alternate state A and B. Whereas the larger movement of cytosine sugar in the state B caused the base to position not only toward the major groove, but also being pulled toward next residue, C(7). The difference in the mode of movement between these two bases results in changes of particularly "shear",

"stagger" and "buckle". As shown in the base-base parameter (figure 2.11-2.12) calculated by the program *CURVES 5.3* (Lavery and Sklenar 1988), the shear is increased by ~0.15Å and the stagger is increased by >0.2Å. The angle of buckle is actually reduced by 10 degrees, causing the G/C base pair in state B to be planer.

The effect of the different modes of movement is also evident in inter-base pair parameters (Figure 2.13-2.14). The values calculated between either one of the alternating states base pairs, G(5)/C(16) or C(6)/G(15) and its adjacent base pair illustrates large deviation in "tilt" and "shift", which both can be explained by the movement of guanine and cytosine toward the major groove in the alternate states, while guanine is rather anchored at C1'. Whereas the small changes in "stretch" and "opening" suggest that the properties of hydrogen bonding interactions between these base pairs are relatively conserved even compared to the original state (A). This appears to be due to the minimized lateral movement of the cytosine residue as it slides to the major groove. The slight lateral movement of the cytosine residue is also compensated by the buckling of the guanine (still anchored at its C1').

Geometry of the sugar-phosphate backbone

Overall, the sugar moieties of the duplex are found to adapt sugar puckers in or neighboring of the C2'-*endo* domain (Table 3.2), except for the O4'-*endo* seen in C(7). The difference between alternate states A and B is subtle. In all cases where multiple states are seen, the puckering either does not differ or the differences are within 36° of the pseudorotation phase angle (Altona and Sundaral.M 1972). The most deviated sugar puckering (from normal C2'-*endo* in B-DNA) is seen in C(7).

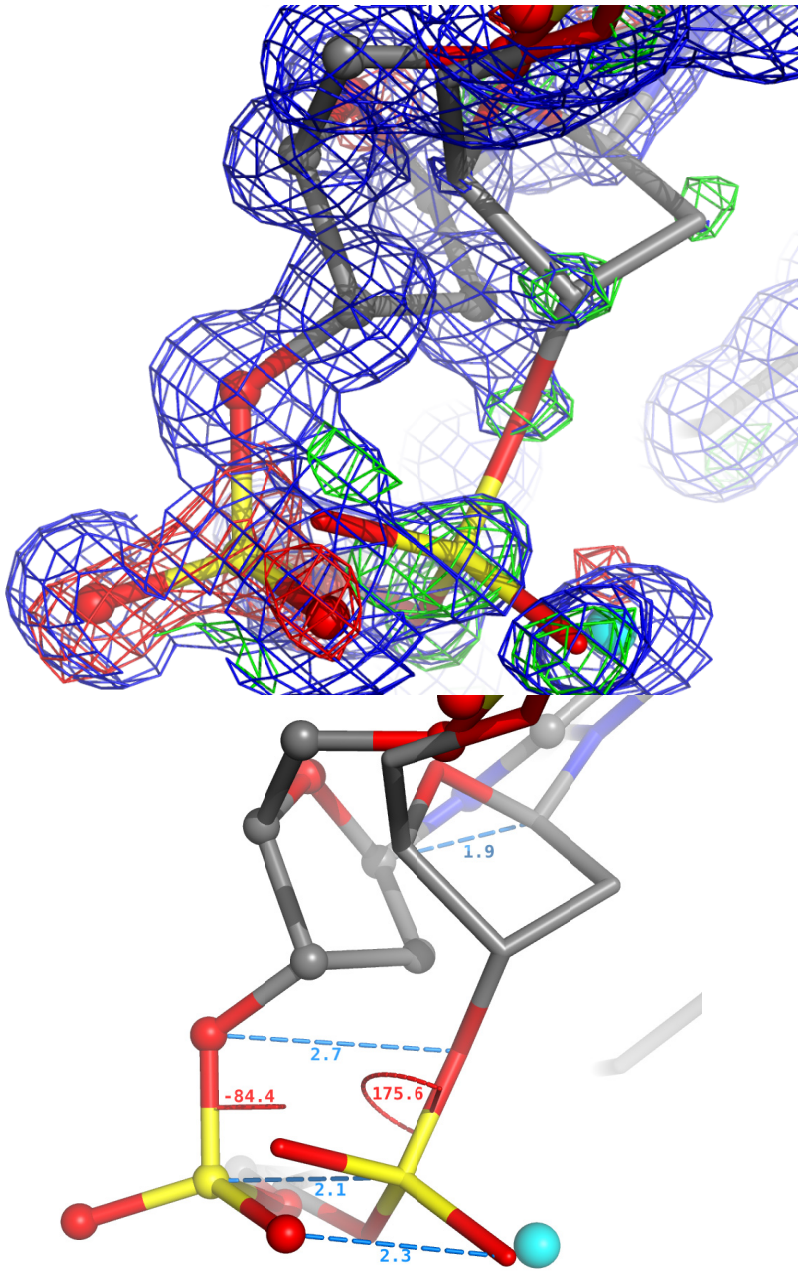

Figure 2.7: Close up of phosphate group at C(6)pC(7) step. In both figures, the original state ("A", 75% occupancy) is displayed in ball and stick model, and it's alternative state ("B", 25% occupancy) is displayed in stick model. DNA coordinates are colored in the CPK coloring scheme, and the neighboring HOH (254A, 75%) is shown in sphere, colored cyan. (Top) Sum electron density ($2F_o$-F_c) after the refinement is contoured at 1σ,colored blue, and both negative and positive difference densities (F_o-F_c) at pre-refinement stage (before the second alternative model was built) is superimposed, contoured at 2.5σ and -2.5σ, colored green and red, respectively. Notice the defined peaks around the original O3'(above phosphorous atom) compared to the elongated peaks around O5' (figure 2.8). (Bottom) Same as left figure minus electron densities. The distances indicating the shift between alternate states are shown in blue. Dihedral angles (C3'-P) are shown in red, indicating B_I-B_{II} phosphate transition at this step as DNA alters its state from A to B (See section "Geometry of the sugar-phosphate backbone").

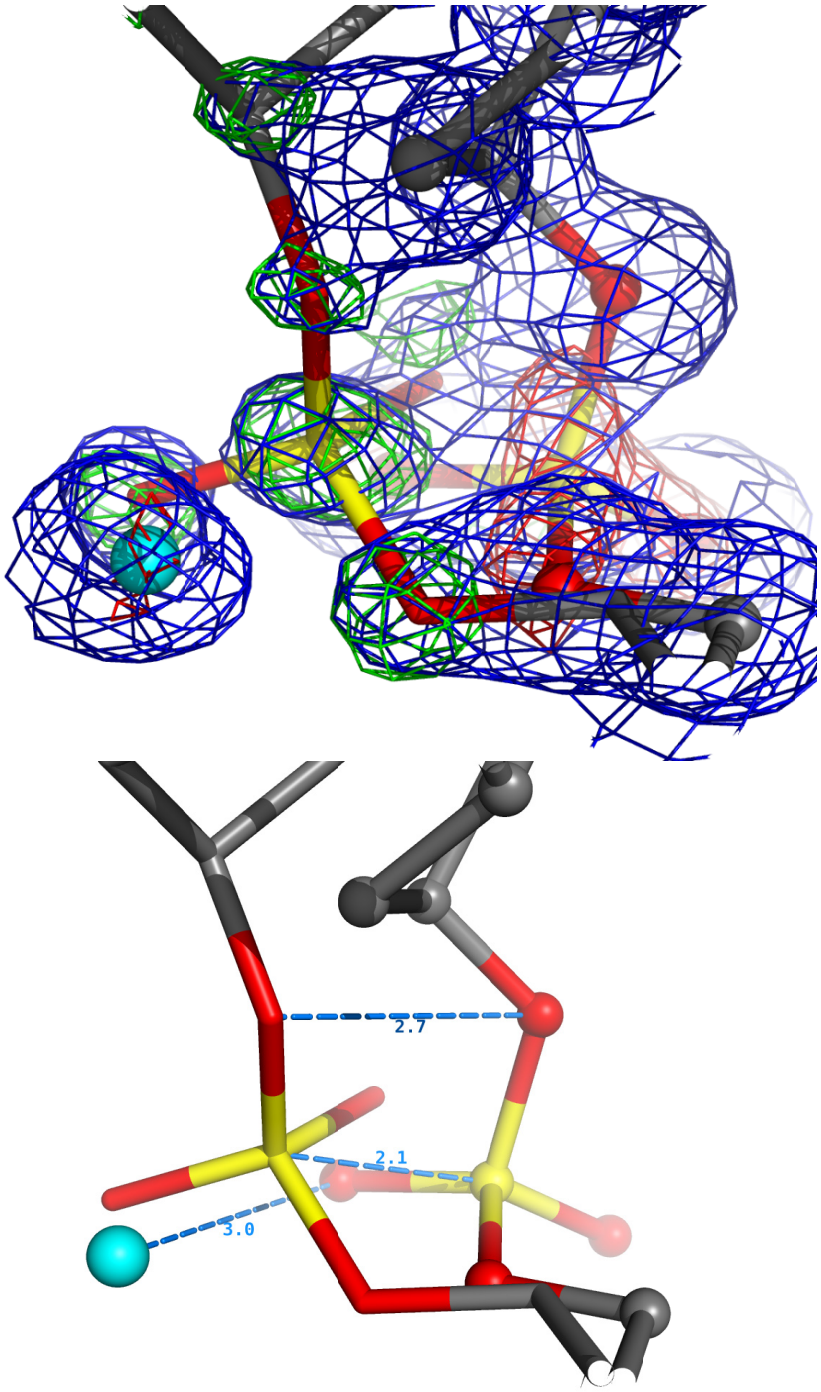

Figure 2.8: Close up of phosphate group at C(6)pC(7) step with 180° turn, respect to the previous figures (figure 2.7). Both figures follow the same scheme as the previous figures. Notice the elongated $2F_o$-F_c peaks next to the original O5' (below phosphorous atom), suggesting the shift/slide of the phosphate group to the left in this angle, as well as the 2 defined peaks around the original C2' (above state B phosphorous atom) explained in the text.

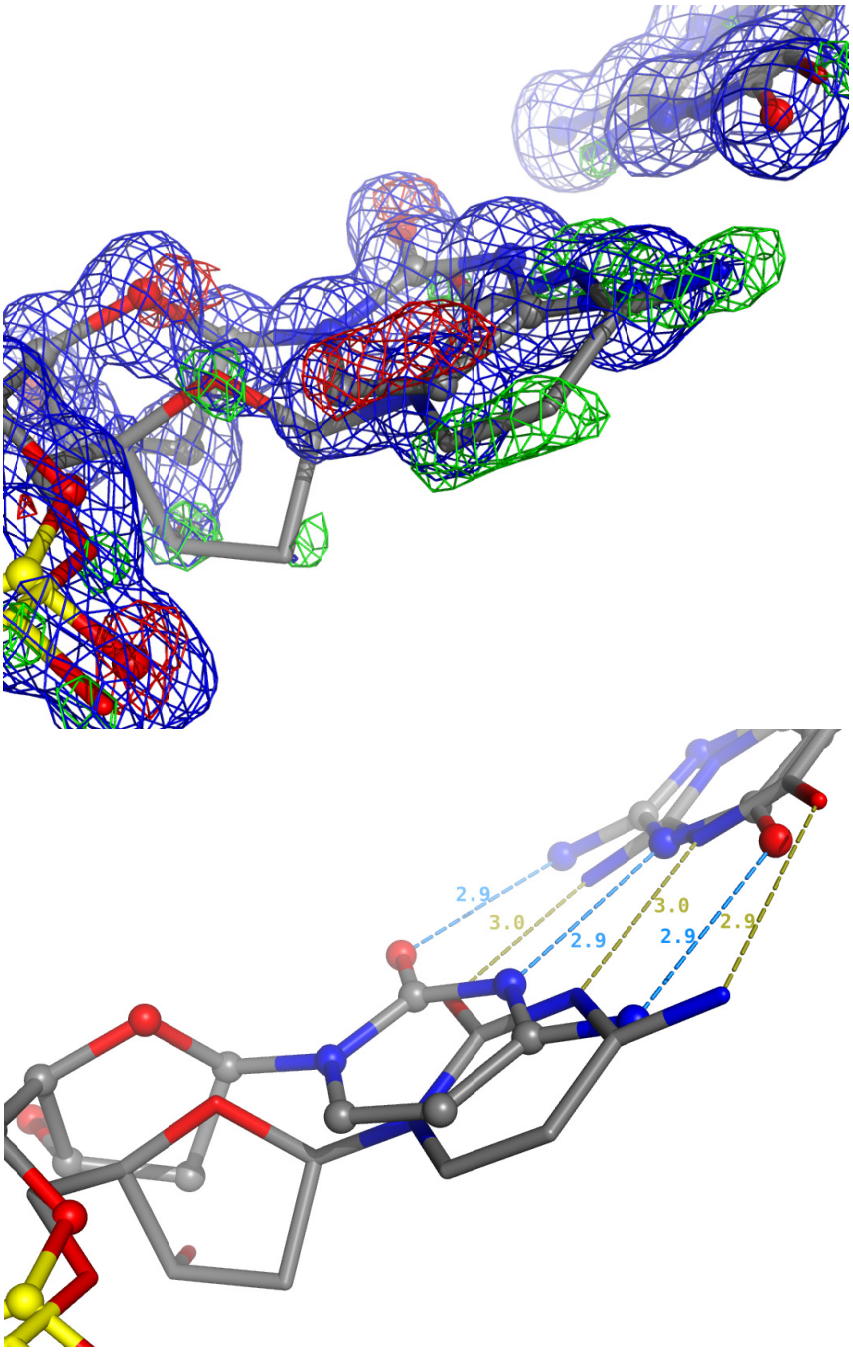

Figure 2.9: Extensive local structural variation of C(6). In both figures, residue C(6) is shown on the left, while its pairing residue, G(15) is shown at the top-right. Original state A is shown with ball and stick, and its alternative state B is shown in stick model. (Top) Shown with refined sum electron density in blue, contoured at 1σ. Pre-refinement (before the alternative model was built) difference electron densities are superimposed, colored green (2.5σ) and red (-2.5σ). Notice the base C(6) is not only shifted toward major groove, but also tilted as transition occurs, with respect to its original model. (Bottom) Without electron densities. Hydrogen bonds distances for the base pairs are shown, blue for state A, and olive for state B.

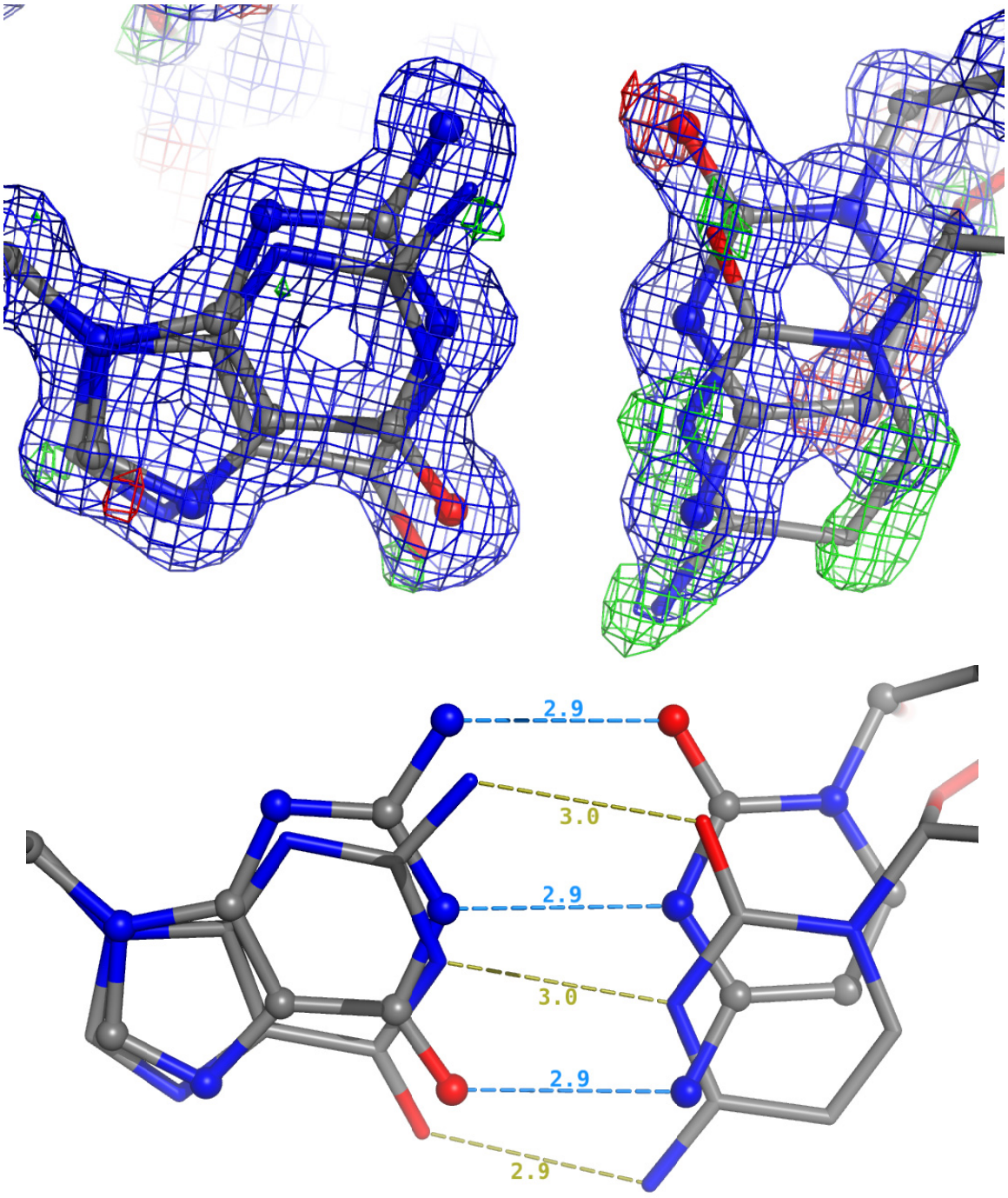

Figure 2.10: Close up views of the base pair G(5)/C(6). Alternative state A (75%) is shown in ball and stick, state B (25%) is shown in stick model. (Top) Shown with sum electron densities (after refinement, 1σ, blue) with difference electron densities (before refinement, 2.5σ/-2.5σ. green/red) superimposed. At this contour level, $2F_o$-F_c peak does not appear at C5 position of cytosine as its peak height is ~0.7σ. (Bottom) Without electron densities. Hydrogen bond distances are shown for the base pair, colored blue for alternative state "A" and Olive for state "B".

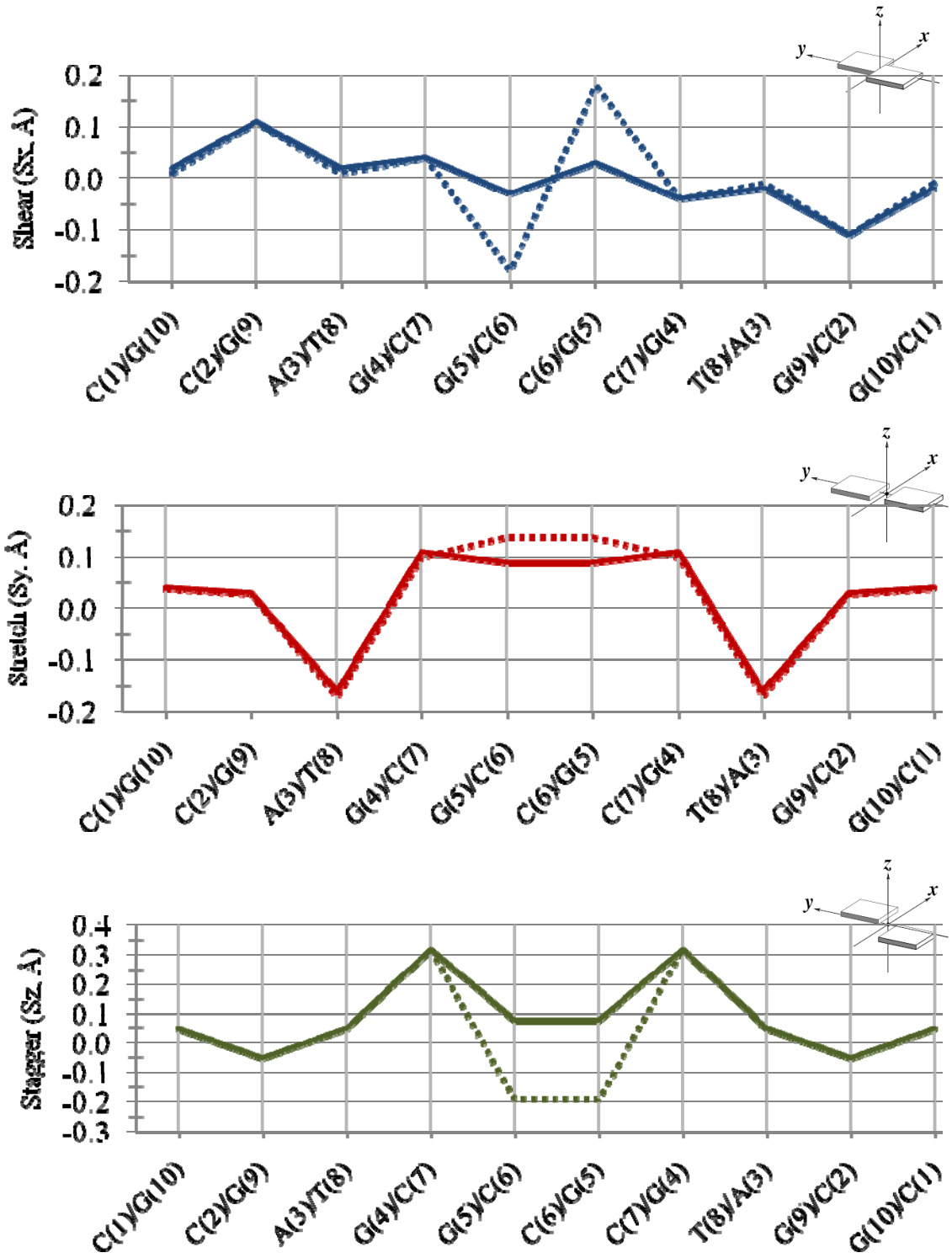

Figure 2.11: Base-base parameters (Shear, Strech and Stagger) as calculated by CURVES 5.3. All plotted in the unit of angstrom (Å) on the y-axis against the x-axis indicating base pairs in the duplex, chain A/B, where the parameters are computed. Solid lines represent alternative state A, whereas dotted lines represent state B. Schematic representation of helical parameters shown in each plot were reproduced from 3DNA (Lu and Olson 2003).

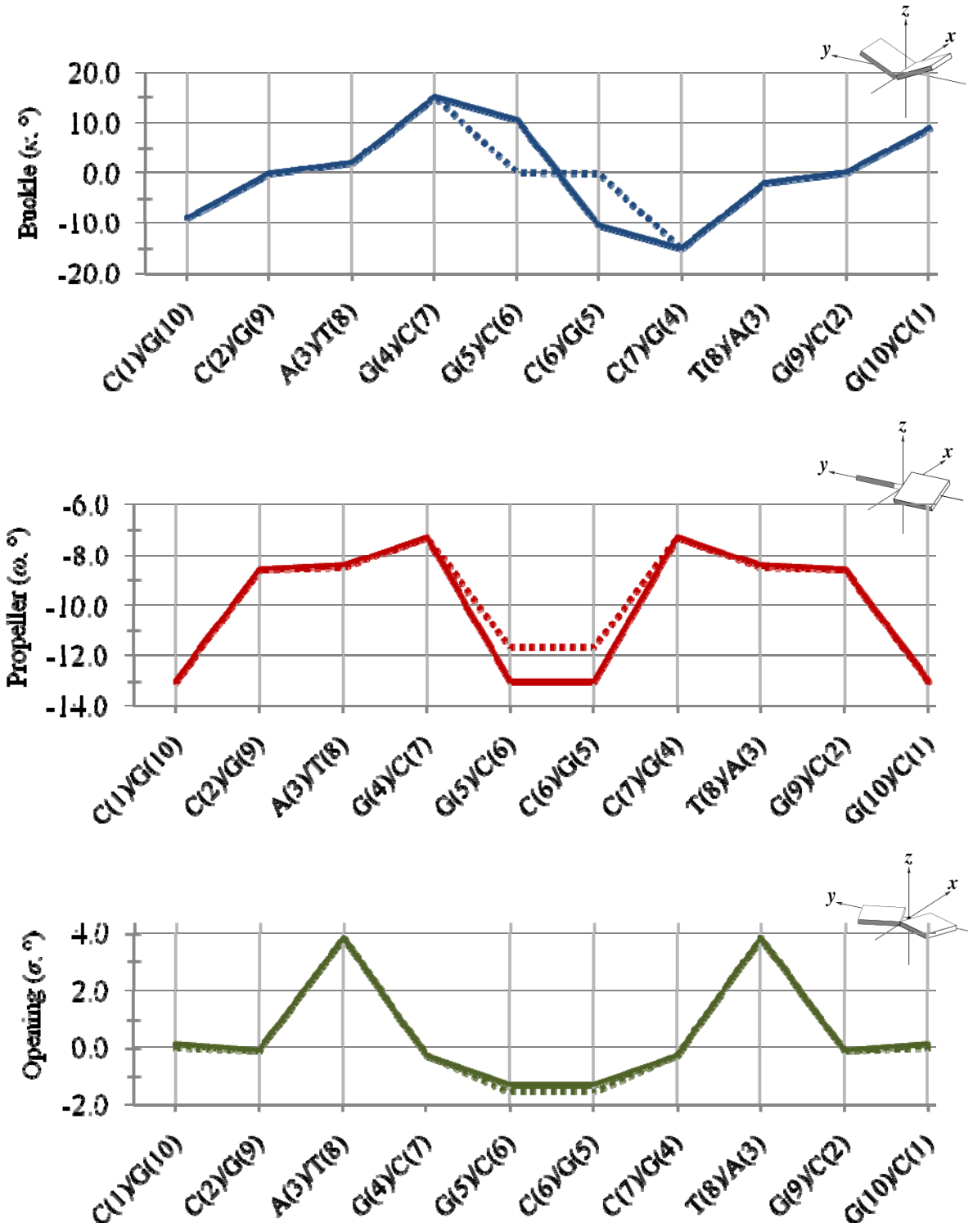

Figure 2.12: Base-base parameters (Buckle, Propeller and Opening) as calculated by CURVES 5.3. All plotted in the unit of degrees (°) on the y-axis against the x-axis indicating base pairs in the duplex, chain A/B, where the parameters are computed. Solid lines represent alternative state A, whereas dotted lines represent state B. Schematic representation of helical parameters shown in each plot were reproduced from 3DNA (Lu and Olson 2003).

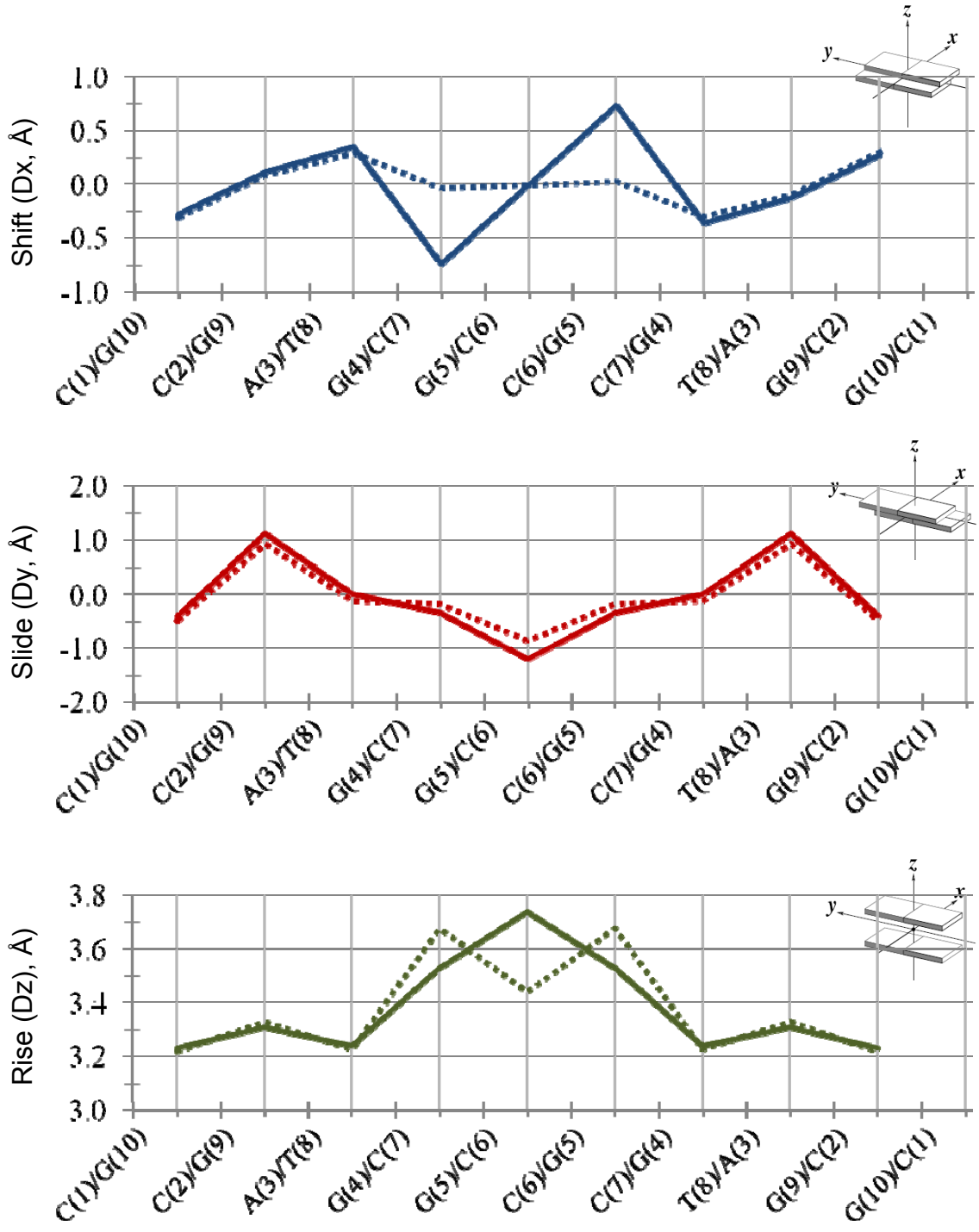

Figure 2.13: Inter-base pair parameters (Shift, Slide and Rise) as calculated by CURVES 5.3. All plotted in the unit of angstroms (Å) on the y-axis against the x-axis indicating base pairs in the duplex, chain A/B. Solid lines represent alternative state A, whereas dotted lines represent state B. Note that the values are computed between two base-pairs shown on x-axis. Schematic representation of helical parameters shown in each plot were reproduced from 3DNA (Lu and Olson 2003).

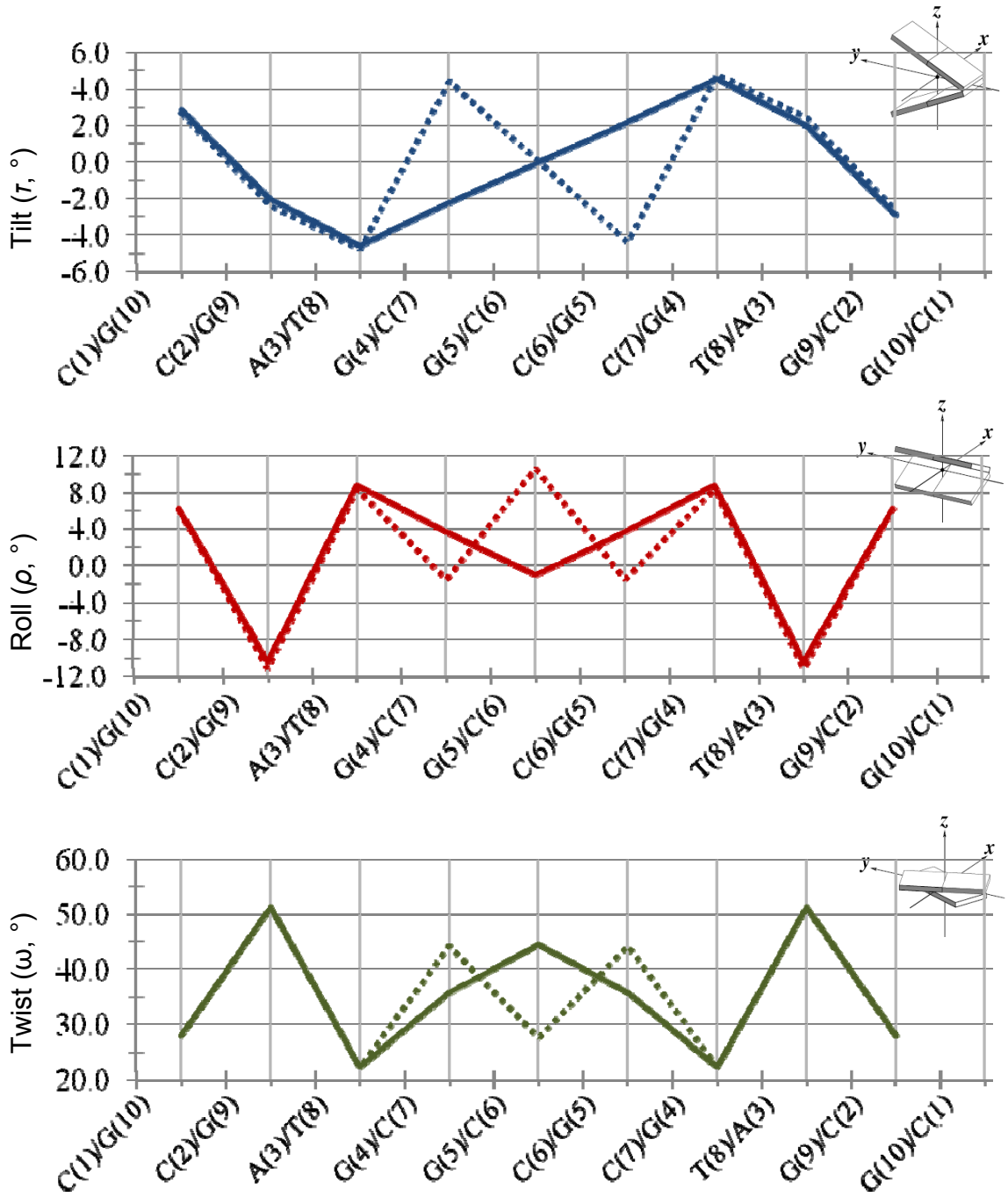

Figure 2.14: Inter-base pair parameters (Tilt, Roll and Twist) as calculated by CURVES 5.3. All plotted in the unit of degrees (°) on the y-axis against the x-axis indicating base pairs in the duplex, chain A/B. Solid lines represent alternative state A, whereas dotted lines represent state B. Note that the values are computed between two base-pairs shown on x-axis. Schematic representation of helical parameters shown in each plot were reproduced from 3DNA (Lu and Olson 2003).

The residue shows O4'-*endo* (A) and C1'-*exo* (B). The mean pseudorotaion phase angle of 150°/149° in state A/B is in accordance with the value of 147° reported for the original structure (Heinemann and Alings 1989). Their individual phase angles are also comparable to either state of the current model or to the averaged phase angle computed from the fractional occupancies of state A & B. For example, their reported phase angle of 109° for C(7) is comparable to 107° (95° * 0.75 + 142° * 0.25) in the current data. Hence, the slight deviation seen in these two is more likely because of their rather averaged final coordinates (vs. defined multiple states) due to the lower resolution.

The anti-correlation of exocyclic tostion angles between ε (epsilon, C4'-C3'-O3'-P) and ζ (zeta, C3'-O3'-P-O5') reported previously by Heinemann and Alings is also observed in the current structure. This anti-correlation creates B-helical subtypes B_I and B_{II}, which influence the position of the phosphate group with respect to the grooves of the double helix (Fratini, Kopka et al. 1982). Subtype B_I is more common in B-DNA, where the phosphate is in a roughly symmetric position with respect to the grooves, while the B_{II} swings around the phosphate towards the minor groove. In some cases, such movement of the B_{II} phosphate pushes the adjacent base into the opposite direction (toward the major groove). This phenomena can also be characterized by the difference in ε - ζ, which passes from roughly -90° in B_I to roughly +90° in B_{II} (Hartmann, Piazzola et al. 1993). Table 2.2 lists the helical subtypes, B_I and B_{II} in the current structure. The torsion angles ε, ζ and the calculated values of ε - ζ are also plotted in the figure 2.15.

The final refined model of $[d(CCAGGCCTGG)]_2$ shows a rather defined alternating pattern of these B_I and B_{II} states, except at rhw C(6)pC(7) step, where the pattern is broken in alternate state A. As shown in the plot of ε - ζ, the B_I-B_{II} transition

occurs at this step, as the state changes from A to B (ε - ζ = -82°→84°). Hence, state B preserves the alternating pattern perfectly throughout the duplex. This results in the helix to conserve the B_{II} subtype in both strands facing one another at every other step. It is interesting to note that the structure presented here reveals the B_I-B_{II} transition at the same site within the helix obtained from a single crystal. The prior studies of B_I-B_{II} transition typically involve surveying multiple x-ray and/or NMR structures of the same (or similar) oligomers (Hartmann, Piazzola et al. 1993; MacKerell 2009).

However, it should be noted that the current description of the effect of B_I-B_{II} transition in terms of overall helical geometry is limited by the assumption that positions of adjacent sugars are relatively conserved. This can be seen from the observation that, while the deviation in torsion angles at the C(6)pC(7) step and the movement of the adjacent base appear to be in accordance with the criteria of B_I-B_{II} transition, the resulting change in the helix is actually opposite to what is commonly thought of the transition. The minor groove is widened at the C(6)pC(7) step, which is apparent in figure 2.16. This is opposite to the effect expected in a typical B_I-B_{II} transition, which swings the phosphorous atom around toward the minor groove, causing the local "kink" into the minor groove. As noted earlier, the variations between these two states extend to the adjacent sugar and the base of C(6), causing both moieties to radically shift toward the major groove. Therefore, the B_I-B_{II} transition at this step is now shadowed and excessively overcome by the pronounced movement because the phosphate is also shifted toward the major groove.

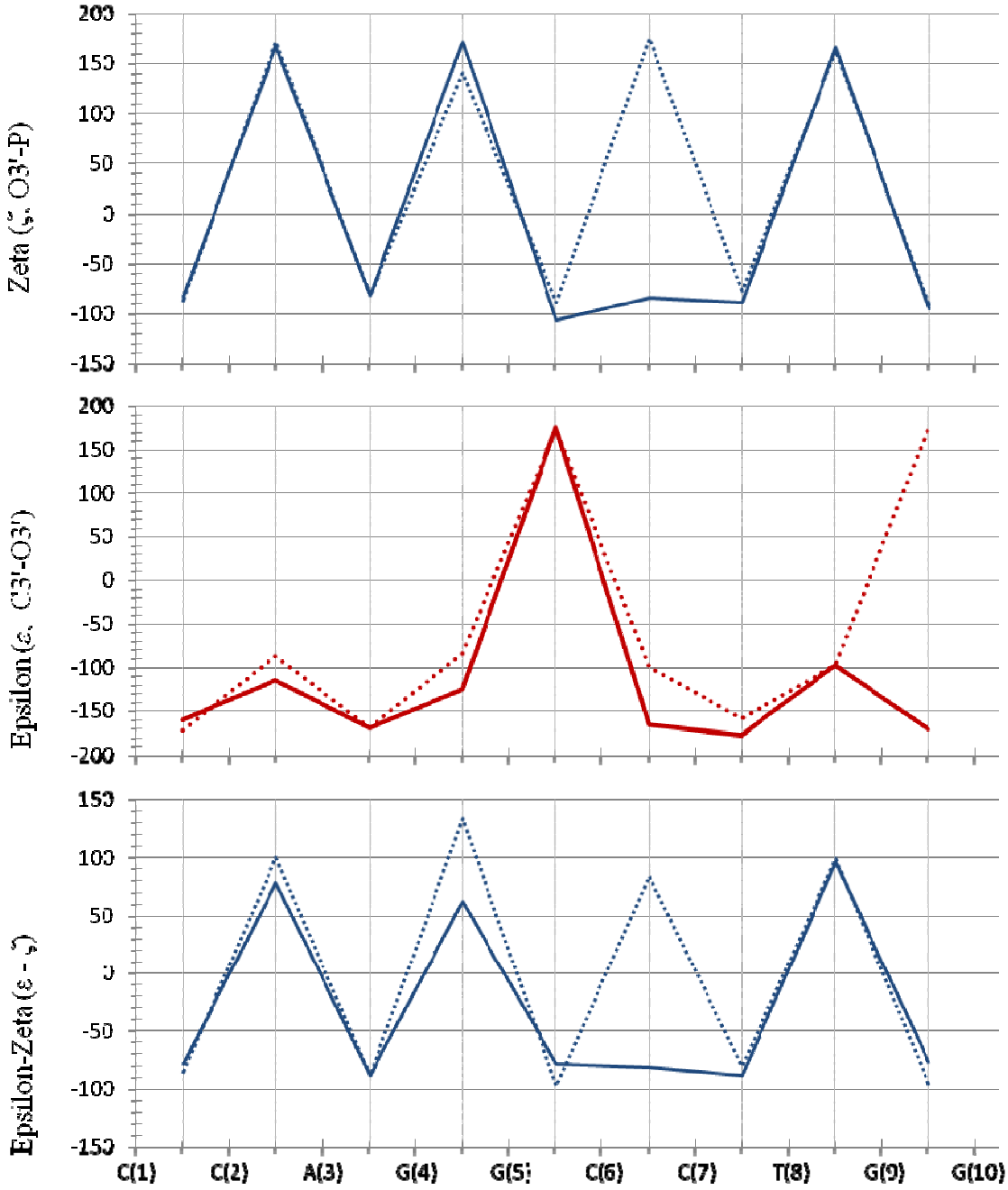

Figure 2.15: Plots of torsion angles as calculated by CURVES 5.3. From the top, O3-P (blue) and P-O5' (red, all plotted in the unit of degrees (°) on y-axis against x-axis indicating base pairs in the duplex. Note that the values are computed between two bases in each strand shown on x-axis. Solid lines represent alternative state A, whereas dotted lines represent state B. The difference in ε-ζ value at C(6)pC(7) step indicates B_I-B_{II} phosphate transition as DNA alters its state from A to B (See section "Geometry of the sugar-phosphate backbone").

Table 2-2: Stereochemistry of the sugar-phosphate backbone

	Main-chain torsion angles (°)															
	α (P-O5')		β (O5'-C5')		γ (C5'-C4')		δ (C4'-P)		ε (C3'-O3')		ζ (O3'-P)		χ (C1'-N)		ε-ζ	
Base	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
C					61	179	125	114	-161	-172	-83	-87	-124	-124	-78	-86
C	-71	-60	174	-179	50	45	140	148	-115	-88	166	172	-87	-87	79	100
A	-65	-85	150	141	50	57	150	135	-170	-170	-81	-81	-84	-84	-89	-89
G	-71	-71	167	167	45	44	134	134	-126	-85	171	141	-82	-82	62	134
G	-65	-70	139	142	50	34	141	131	175	173	-106	-89	-113	-101	-79	-98
C	-61	-62	-175	-175	56	48	137	146	-166	-100	-84	176	-116	-96	-82	84
C	-65	-69	166	133	50	56	98	122	-178	-159	-89	-78	-122	-122	-89	-82
T	-61	-77	173	170	50	47	142	129	-99	-99	165	162	-91	-91	96	99
G	-73	-73	153	153	50	42	146	141	-172	173	-95	-91	-80	-80	-77	-96
G	-64	-60	164	174	47	49	123	135					-97	-97		
Mean	-66	-70	164	121	51	60	133	134	152	139	124	135	100	-96		

	Helical subtype		Amplitude (τ_m, °)		Phase (°)		Sugar Pucker	
Base	A	B	A	B	A	B	A	B
C	B_I	B_I	28	26	145	117	C2'-endo	C1'-exo
C	B_{II}	B_{II}	43	43	153	153	C2'-endo	C2'-endo
A	B_I	B_I	32	32	181	181	C3'-exo	C3'-exo
G	B_{II}	B_{II}	45	49	135	136	C1'-exo	C1'-exo
G	B_I	B_I	41	25	155	159	C2'-endo	C2'-endo
C	**B_I**	**B_{II}**	34	46	166	156	C2'-endo	C2'-endo
C	B_I	B_I	44	31	95	142	O4'-endo	C1'-exo
T	B_{II}	B_{II}	44	45	150	135	C2'-endo	C1'-exo
G	B_I	B_I	36	27	196	172	C3'-exo	C2'-endo
G			41	38	123	144	C1'-exo	C2'-endo
Mean	B_{II}	B_{II}	39	36	150	149	C2'-endo	C2'-endo

Values are shown for both alternate states A & B. The torsion angles, pseudorotation parameters, τ_M=amplitude of pucker, P=phase angle of pseudorotation (Altona and Sundaral.M 1972) and sugar pucker were determined by CURVES 5.3 (Lavery and Sklenar 1989). The helical subtypes, B_I and B_{II} are determined from the values of the torsion angles ε and ζ as described in the text. Note that torsion angles were normalized into one direction prior to determining mean values and ε - ζ. See also figure 2.15 for the plots of torsion angles ε and ζ.

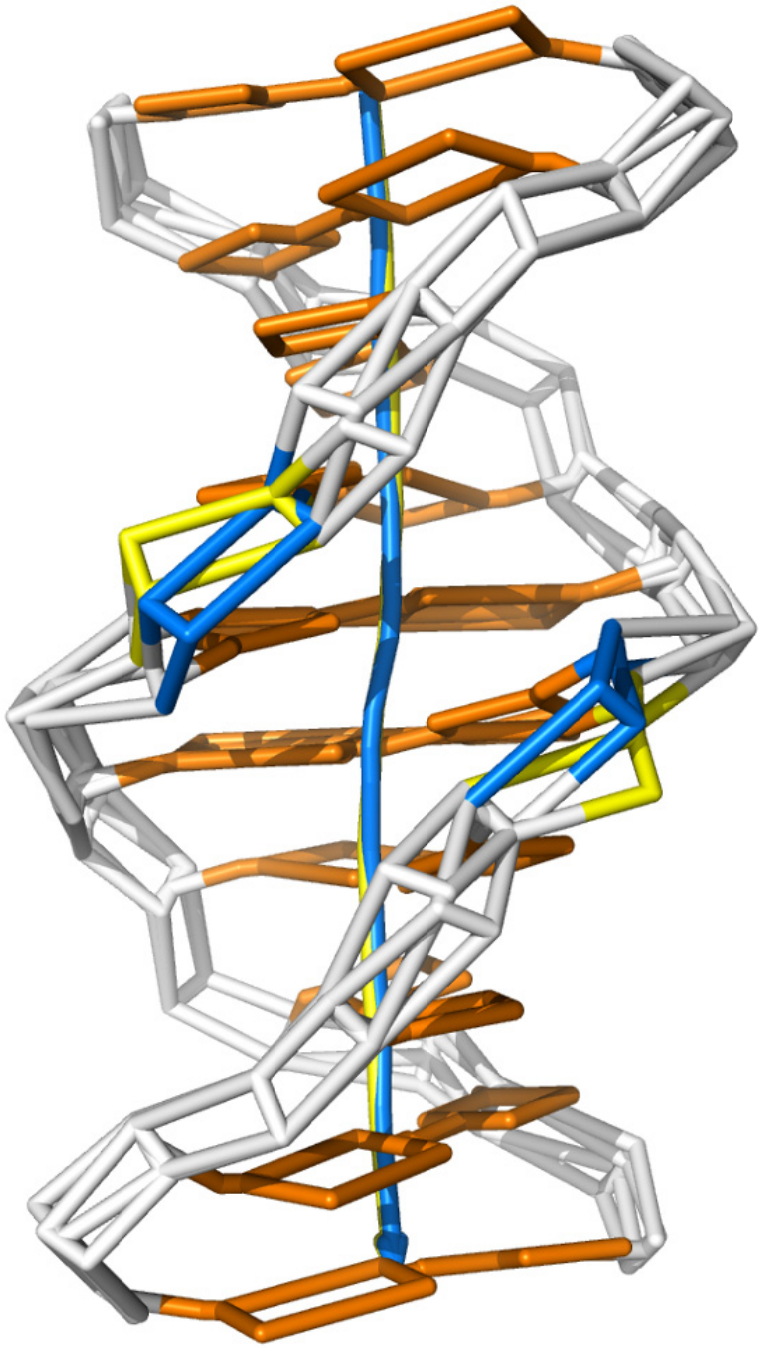

Figure 2.16: Schematic representation of [d(CCAGGCCTGG)]$_2$. Bases are colored orange and the ribs of backbones are colored white, except at the C(6)pC(7) step in both strand; state A is colored blue and its alternative state B is colored yellow at these steps. Notice the widening of the minor groove as state transition occurs from A to B, which is opposite to the effect seen in a typical B_I-B_{II} transition of phosphate groups (see text). Helical axes through the center of duplex are also shown for both DNA state A (blue) and state B (yellow). The figure was constructed by CURVES 5.3 (Lavery and Sklenar 1989) and rendered with Pymol v1.1 (DeLano).

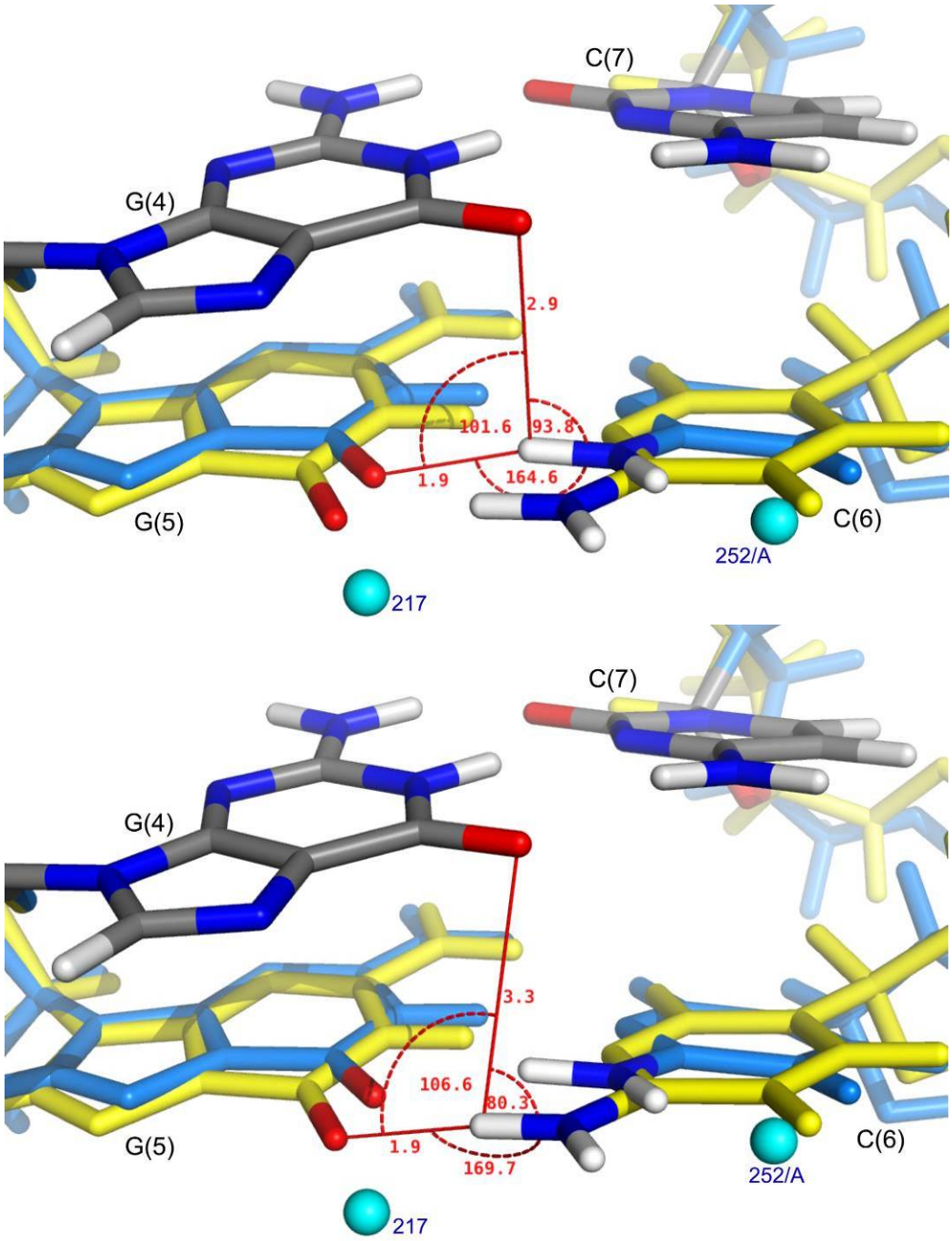

Figure 2.17: Three-center (bifurcated) H-bond at C(6)C(7)/G(4)G(5) step observed. ***Related angles and distances in each state model are shown in (above) state A, (below) state B***. *In both figures, DNA bases with multi conformations are colored with (light blue) state A and (yellow) state B, except for the atoms involved in angle/distant measurements. Hydrogen atoms are placed by the program Molprobity to be coplanar with the attached residues (Davis, Leaver-Fay et al. 2007).Neighboring water molecules, 252 and 217 are shown in sphere, colored cyan for a reference purpose. Note that the occupancy of HOH217 is 100%, whereas HOH252 (75%) only exists when nearby base C(6) is in state A. DNA coordinates, water/ions that are not involved are omitted for clarity.*

Three-center (bifurcated) hydrogen-bonds

The pronounced local variations displayed by the two alternating states of the residue C(6) appear to affect another aspect of local helical geometry: the formation of the major groove inter-stand three center (bifurcated) H-bonds (Figure 2.17). Inter-strand bifurcated H-bonds are frequently observed at the ApA step of the A-tracts (Coll, Frederick et al. 1987; Nelson, Finch et al. 1987). In an A-tract, the narrow minor groove and large propeller twist of the base pair permits the N6 of adenine to form H-bonds with the O4 atoms of the two adjacent thymine residues. It has been suggested that the existence of bifurcated H-bonds (along with abundant occurrence of interstrand amino contacts in CpG and ApT steps in B-DNA) is facilitated by partial pyramidization (sp^3) of the amino groups (Sponer, Leszczynski et al. 2001).

At the C(6)C(7)/G(14)G(15) step of the current structure, both alternate models show the contact distances between the N4 atom of C(6) and the O6 atoms of two adjacent guanine residues to be identical (3.2Å to 5' side G and 2.9Å to the opposing G). However, the geometries of these H-bonds differ between the 2 states. In definition, all 3 angles by centering $H4_2$ should be greater than or equal to 90°. The coplanarity of the 4 atoms involved (N4 of cytosine, covalently linked H and two O6 atoms in adjacent guanine residues) should also be indicated by the sum of three angles to be close to 360° (Jeffrey and Mitra 1984). While both states meet the latter part of criteria (360°/357° in states A/B), the major deviation is seen in the angle between N4-$H4_2$-O6/G(4), as state A displays an angle of 94°, the same angle is reduced to 80° in state B model. The distance between $H4_2$-O6/G(4) is also extended to 3.3 Å (vs. 2.9Å) in state B. The state A thus

clearly exhibits inter-strand three-center (bifurcated) H-bonds, while state B appears to fall short to meet the criteria.

The pronounced shift of the C(6) base does not affect the base pairing H-bonds, as its opposing guanine is also shifted/rotated. However, the propeller twist at the C(6)/G(15) base pair is slightly reduced as the A-B state transition occurs (-13.02° → -11.64°), and more pronounced change is seen in the buckle (-10.55° → -0.13°). These changes take place while the position of adjacent G(4) is conserved in a single state, hence creating geometric deviation involving bifurcated H-bonds as the A-B state transition occur at C(6). The original structure by Heinemann and Alings showed the angle between N4-H4$_2$-O6/G(4) being <90° (86°) at the same CC/GG step. This led the authors to conclude they had observed a "boaderline case of bifurcated H-bonding". This is again, more likely the case of their averaged base position, which is more accurately a discrete disordered region with two alternating states as shown in this work.

Water and counter-ions

Multiple discrete states of Mg^{2+}

There are a total of five partially occupied hexaaqua-Mg^{2+} ions per asymmetric unit (per single strand or equivalently, per half of the duplex). The Mg^{2+} complex is numbered 21-25, accompanied by the letter designation (A-C) to specify alternative states, where necessary. Each Mg^{2+} complex contains six water molecules coordinating to a metal center, and they are arbitrarily named O1-O6. None of the Mg^{2+} complexes are fully occupied, with the estimated occupancies ranging from 15% to 90%. Some Mg^{2+} complexes appear to occupy two or three alternative states, indicated by a hydration shell

overlap. Mg(22) occupies either state A or state B, with an estimated occupancy of 60% (A) and 15% (B). Mg(24) occupies three states, A, B and C, with an estimated occupancy of 35% (A), 15% (B) and 35% (C) (figure 2.18-3.19). The sum of the occupancies is always less than 100%. Other Mg^{2+} complexes access single states. Mg(21) is the most highly occupied complex of all at 90%. Mg(23) has 40% occupancy with letter designation A, since it shares the site with partially occupied water molecules. Mg(25) occupies a crystallographic special position, where a 2-fold symmetry axis directly runs through its metal center; hence its occupancy of 40% should be doubled when considering a duplex, giving a true occupancy of 80%. A summary of the multiple states model, occupancies, thermal factors and average bond lengths is given in table 3.3.

Disordered Mg^{2+} and multi states of DNA backbones

The interactions of Mg^{2+} complexes with the DNA backbone involve hydrogen bonds between the first shell water molecules and the oxygen atoms in the sugar/phosphate backbones, O1P, O2P, O3' or O4'. Interactions between the O5' and Mg^{2+} complex are not seen in the current structure. In some cases, both phosphate oxygens are bridged by first shell water molecule(s) on the same Mg^{2+}. As seen in figure 2.20 and table 3.4, most phosphate groups make contact with the first shell water molecules of one or two Mg^{2+} complexes. In total, the interactions enable lateral packing of helices in the crystal lattices through charge neutralization and hydrogen bonding. However, the local electrostatic environment surrounding each phosphate group is expected to vary as Mg^{2+} ions are found to present disordered and/or partially occupied in the current structure. Therefore, their relative positions with respect to each phosphate group may in turn correlate with the frequently appearing multi-states of the DNA

backbones. Overall, the relationship between the multiple states seen in the DNA backbones and partially occupied/disordered Mg^{2+} ions in the current structure can be categorized into the following two.

Single cation Bound/Free mode

Single state Mg^{2+} complexes appear to affect the position of DNA phosphate groups due to its partial occupancies. This mode of cation/DNA correlation is observed when there is a partially occupied Mg^{2+} ion interacting with a phosphate group. This correlation causes phosphate groups to be disordered, including both the G(4)pG(5) and the C(7)pT(8) steps (figure 2.21a, 2.21b). The occupancies of the bound Mg^{2+} complexes at these steps are estimated to be 40-60%, creating two discrete electrostatic environments. These sites are described here as bound and free. For a given phosphate group the bound state is directly associated with the Mg^{2+} complex through hydrogen bonding interactions with the first shell water molecules. The free state of the DNA either does not engage in hydrogen bonding interactions with first shell water molecules or the number of interactions is found to be less. At G(4)pG(5) step, the phosphate in state A forms hydrogen bonding interaction with Mg(22) at distance of 2.6Å (bound state), whereas the interaction with the phosphate in state B is not feasible as the contact distance becomes too short at 2.3Å (free state). On the other hand, the position of Mg(23) does not conflict with either of the alternative states occupied by the phosphate group at the C(7)pT(8) step. However, the number of hydrogen bonding interactions differs as the phosphate takes alternative states. The phosphate in state A forms 2 hydrogen bonding interactions with first shell water molecules of Mg(23), whereas only one interaction is possible when in state B. Therefore state A is assumed to be the bound state.

Table 2.3: B-values and bond length in magnesium complexes

Residue	Occu.	Metal	O1	O2	O3	O4	O5	O6	b-value	d_{mean} (Å)
21	0.90	9.31	12.88	9.48	11.98	9.22	11.45	10.44	10.68	2.15
22a	0.60	9.81	9.35	9.80	10.52	9.48	9.28	9.81	9.72	2.17
b	0.15	7.03	7.38	8.10	8.01	7.63	6.84	7.73	7.53	2.18
23a	0.40	7.96	8.33	9.04	8.16	7.92	9.00	9.57	8.57	2.17
24a	0.35	11.37	10.04	10.21	10.17	10.43	10.41	11.05	10.53	2.17
b	0.15	6.68	6.69	5.81	6.45	6.58	6.81	6.93	6.56	2.17
c	0.35	11.4	11.24	12.26	11.74	12.3	12.36	12.76	12.01	2.18
25	0.40[1]	6.66	6.43	6.11	5.87	8.94	8.29	7.97	7.18	2.16

The column "Residue" indicates residue number for the hexacoordinated magnesium ions, as listed in the deposited PDB file, followed by letter designation for its alternate state, if any. Occu.: Crystallographic occupancy for the complex. Note that each complex was refined as a "whole", hence all atoms in each complex have the same occupancies. Last 2 columns show mean values. B-value: Mean Isotropic B-value (thermal factors), d: Mean distance between magnesium metal center and each coordinated water, O1, O2, O3, O4, O5 and O6 in the unit of angstrom.

[1]Magnesium complex 25 is located on a crystallographic special position (directly on the 2-fold symmetry axis); therefore its true occupancy is 0.8.

Anchored Bound/Free mode

When there is more than one Mg^{2+} complex near a disordered phosphate, one of the Mg^{2+} complexes appears to anchor the relative positions of the phosphate, while the other appears to correlate with the positional change of the phosphate more exclusively. This mode of cation/DNA correlation is observed at C(1)pC(2)pA(3) and G(5)pC(6)pC(7) steps (figure 2.21c-2.21e). At each step, one Mg^{2+} forms contacts with phosphate oxygens regardless of the DNA state (anchoring); however the second Mg^{2+} can form an extra contact with only one state of the phosphate. Therefore, for a given phosphate group the bound state is directly associated with the second Mg^{2+} complex through hydrogen bonding interactions with the first shell water molecules. Interaction

between the free state of the DNA and the second Mg^{2+} complex is not observed. While the effect of the second Mg^{2+} is similar to what is observed in the former mode of cation/phosphate correlation, the positions of the second Mg^{2+} relative to the phosphate group can be further away. This is seen at C(1)pC(2) and C(6)pC(7) steps, where the two distant Mg^{2+} ions are flanking the phosphate from both sides of the grooves. Hence, the second Mg^{2+} ions at these steps form contacts with O3' or O4' atoms of the backbone instead of O1P or O2P.

There appears to be no correlation with occupancy of Mg^{2+} complexes, as to which Mg^{2+} anchors or induces the change in the phosphate position. However, the preferred state of phosphate group (state A, higher occupancy) appears to form contacts with both anchoring and the second Mg^{2+} ions, whereas only anchoring Mg^{2+} is involved in H-bond contacts when the phosphate is in less preferred state (state B). The sum of the occupancies among contacting Mg^{2+} is always higher when phosphate is in the preferred state, A. This, in turn, may also indicate that the occurrence of the second Mg^{2+} stabilizes the preferred phosphate state. Whereas the second Mg^{2+} either has no effect on the backbone when in a less preferred state or it does not allow phosphate to be in that state, as observed in the case of G(5)pC(6) step because of a too-close contact (2.3Å).

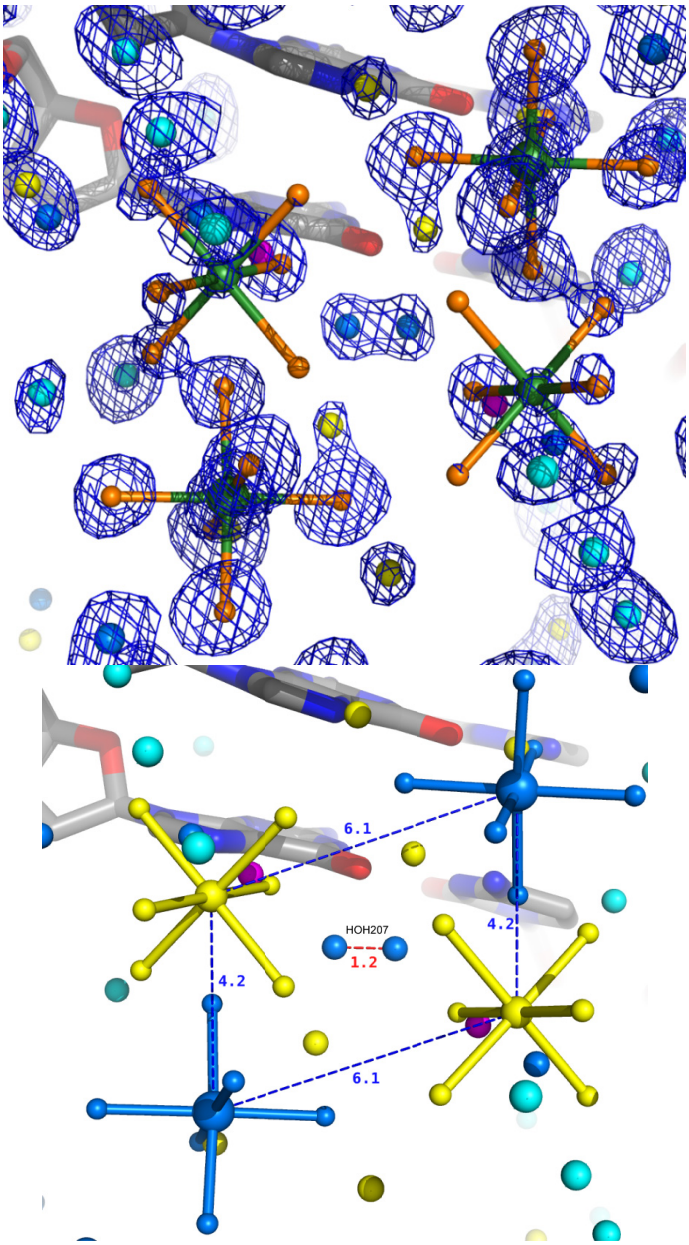

Figure 2.18: Disordered Mg^{2+} complex (Mg22) with 2 alternative states A and B, shown with another set of same residue after 2-fold symmetry operation (-X, Y, -Z) with translation (0 0 -1). One conformation in before and after the symmetry operation makes at least one contact with neighboring base pair G(10)/C(1). Total occupancy among 2 multiple alternating states is 75%, while there are additional water molecules assigned as 3rd conformation, C. Water molecule 207 is slightly off centered from the 2-fold symmetry axis (at the center of each figure), resulting its assigned letter conformation "A" to be rather arbitrary, since both water molecules (before and after the symmetry operation) cannot co-exist at the same time, indicated by the distance of 1.2Å. Hence, as described in the text, the molecule has been treated as special position molecule, with the occupancy halved. (Top) Mg center is colored green, coordinated water molecules are colored orange.Neighboring water molecules are colored cyan including the ones within the disordered complex. Sum electron density is contoured at 1.0σ displayed with mesh, colored blue. Sum density is shown only around magnesium complexes and water for clarity. (Bottom) Same as the top figure without $2F_o$-F_c densities. Mg complexes and water molecules within the disordered region are colored according to the assigned conformation, A=blue, B=yellow and C=magenta. Neighboring water molecules that are not part of the disordered complex are colored cyan. The distances between each conformation taken by Mg ion are also indicated (between metal centers).

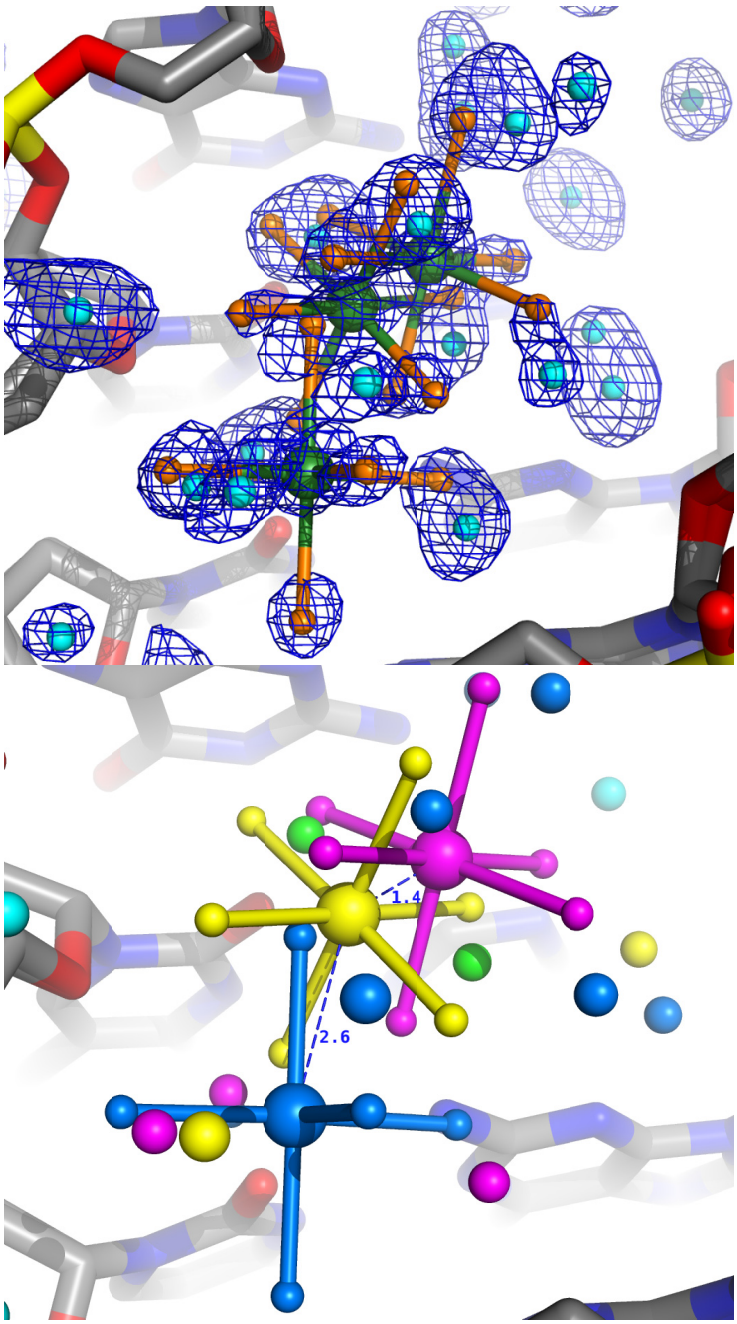

Figure 2.19: Minor-groove bound Mg^{2+} complex (Mg24) with 3 alternative states, A, B and C. Total occupancy among 3 multiple alternating states is 90%, while there are additional water molecules assigned as 4th conformation, D. The disordered complex appears near A(3)pG(4) step in the minor groove. Hydrated magnesium ions are shown with ball and stick model, and neighboring water molecules are shown in small sphere. (Left) Mg center is colored green, coordinated water molecules are colored orange. Neighboring water molecules are colored cyan including the ones within the disordered complex. Sum electron density is contoured at 1.0σ displayed with mesh, colored blue. Sum density is shown only around magnesium complexes and water for clarity. (Right) Same as left without $2F_o$-F_c densities. Mg complexes and water molecules within the disordered region are colored according to the conformation, A=blue, B=yellow, C=magenta and D=green. Neighboring water molecules that are not part of the disordered complex are colored cyan. The distances between each conformation taken by Mg ion are also indicated (between metal centers).

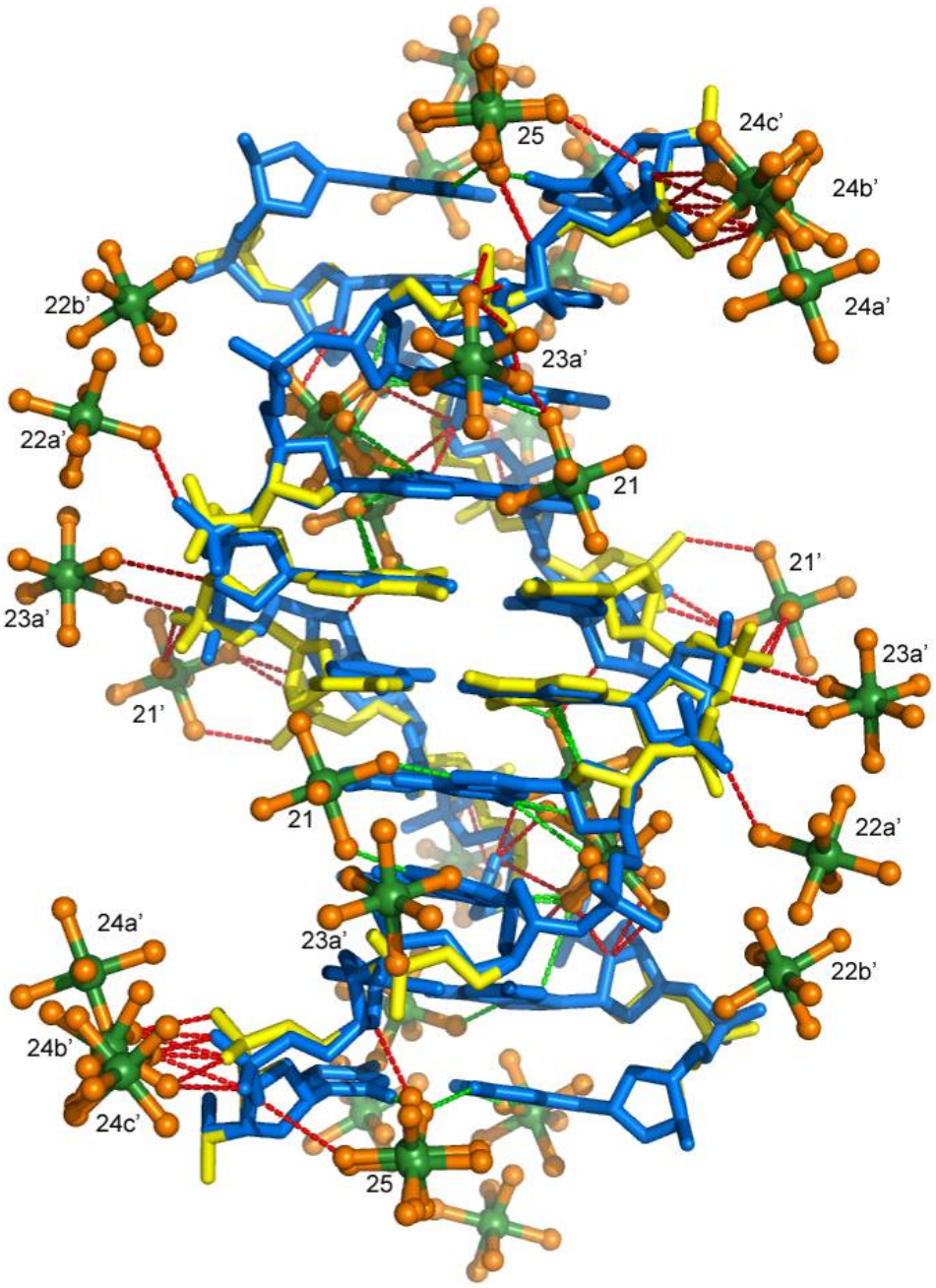

Figure 2.20a: Mg^{2+} within H-bond contacts of $[d(CCAGGCCTGG)]_2$, viewing into major groove. See figure 2.20b for the description of the figure.

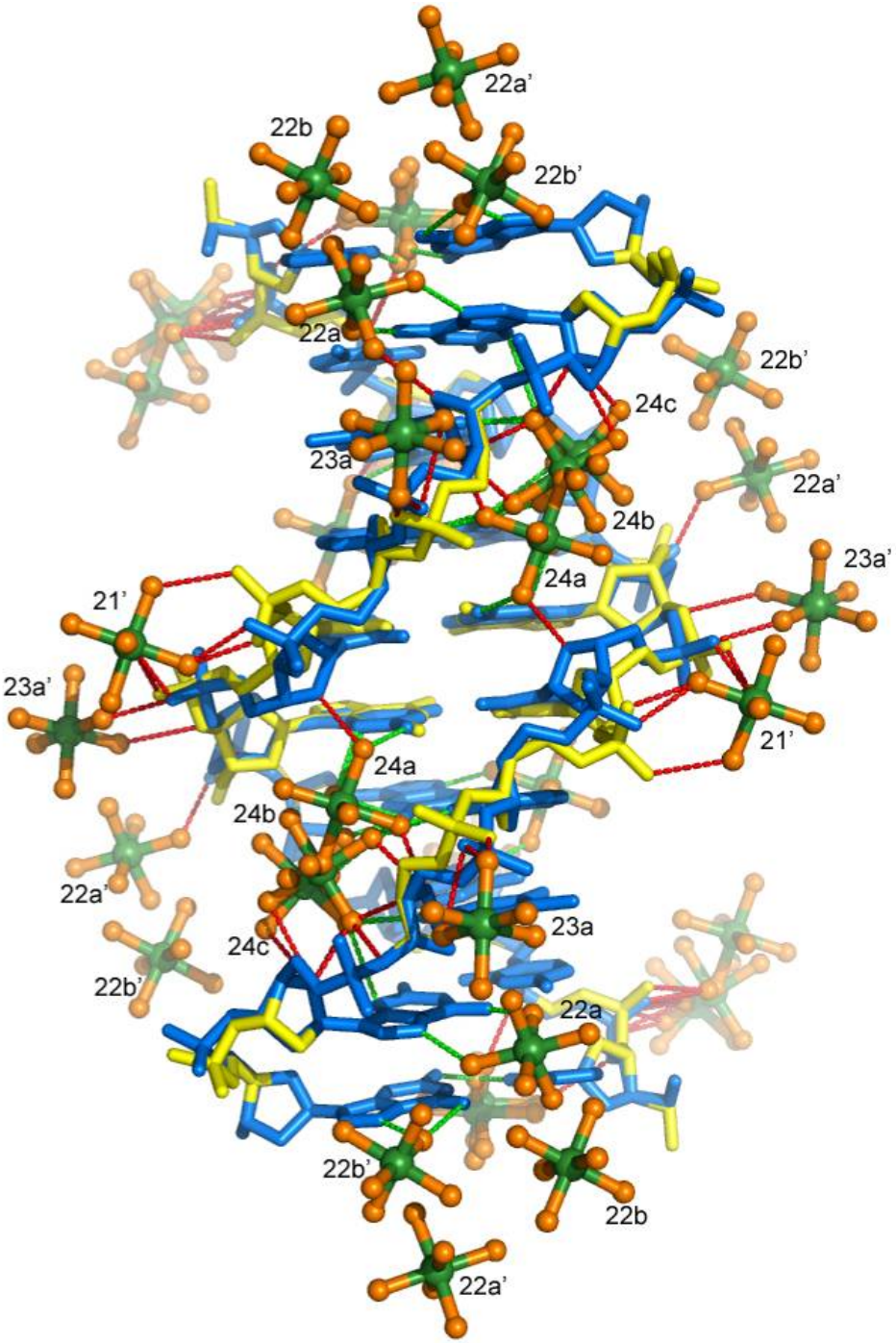

Figure 2.20b: Mg^{2+} within H-bond contacts of $[d(CCAGGCCTGG)]_2$, viewing into the minor groove. DNA is shown with stick model, colored blue while regions with alternate states (state. B) are colored yellow. Each hexaaqua Mg ion is shown with ball and stick, colored green at its metal center and orange for the coordinated water molecules. Contacts between Mg complexes and sugar/phosphate backbones are indicated by red dashed-lines, while contacts with atoms in the bases are indicated with green dashed-line. Each Mg complex is numbered as shown in the coordinate file, followed by a letter designation for the conformation it takes, if any. Prime (') after the residue identifier indicates the symmetry related ion complex. Note that all alternating states within the same residue appear on the figure, if any of the alternating states makes contacts with DNA. Contacts between mis-matched alternating states, that are not theoretically feasible (<2.5Å) are not included in the figure.

Table 2.4: Sugar/Phosphate Backbones -Magnesium interactions

C(1)	O3'a-O2/Mg25[a] (3.43)			
C(2)	OP1a-O3/Mg24b[b] (2.92)	OP1a-O3/Mg24c[b] (3.12)	OP1b-O3/Mg24b[b] (3.33)	OP2a-02/Mg24b[b] (2.90)
	OP2a-O3/Mg24c[b] (3.24)	OP2b-O2/Mg24b[b] (2.84)	OP2b-O3/Mg24b[b] (3.21)	
	OP2b-O3/Mg24c[b] (2.62)	O4'-O5/Mg25 [a] (2.88)	O4'-O4/Mg25 [a,c] (3.47)	
A(3)	OP1a-O2/Mg21 (2.70)	OP1a-O4/Mg23a[d] (2.64)	OP1b-O4/Mg23a[d] (2.63)	OP1b-O1/Mg23a[d] (2.92)
	OP2a-O1/Mg23a[d] (2.75)	OP2b-O1/Mg23a[d] (3.40)		
G(4)				
G(5)	OP2a-O6/Mg22a[e] (2.61)	O3'b-O6/Mg23a[e] (3.54)		
C(6)	OP2a-O5/Mg21[f] (2.71)	OP2b-O5/Mg21[f] (2.57)	OP2a-O4/Mg23a[e] (2.70)	O4'a-O5/Mg24a (2.79)
C(7)	OP1a-O4/Mg21[f] (2.82)	OP1b-O1/Mg21[f] (3.05)	OP2b-O4/Mg21[f] (2.65)	
T(8)	OP1b-O2/Mg23a (2.72)	OP2a-O2/Mg23a (2.50)	OP2a-O3/Mg23a (3.35)	O4'-O6/Mg24b (3.33)
	O4'-O4/Mg24a (2.85)	O4'-O1/Mg24b (3.22)	O3'-O6/Mg24b (3.48)	
G(9)	O4'-O6/Mg24b (2.63)	O4'-O5/Mg24c (3.24)	O4'-O2/Mg24b (3.4)	O4'-O2/Mg24c (2.75)
G(10)				

Contacts between Sugar-Phosphate backbones and Mg^{2+} complexes are tabulated. Left most column indicates the "residue (residue number)". In the table, numbers in parenthesis display distance between the selected atoms in angstrom (Å). Small letters, a, b & c after each atom notation in the table indicates alternative state. First entry: O3'a-O2/Mg25 means the interaction between O3' in state A and 2nd coordinated water of Mg^{2+} complex 25. Only asymmetric unit of DNA (single strand) is included. The partner strand is related by a 2-fold axis with a symmetry operation (-X, Y, -Z) and contacts are identical. All surrounding Mg^{2+} complexes that appear within 3.5Å are included, including symmetry related ones. Super-scripted letters on the tabulated magnesium complexes indicates symmetry operations as listed below.

[a] *Magnesium complex on a Special Position (2-fold related)*
[b] *X-1/2, Y+1/2, Z*
[c] *-X, Y, -Z*
[d] *-X, Y+1, -Z*
[e] *-X+1/2, Y+1/2, -Z*
[f] *X+1/2, Y-1/2, Z*
[g] *X, Y, Z+1*

Table 2.5: Base-Magnesium interactions

C(1)	N4-O2/Mg22a (3.5) **O2-O4/Mg25[a] (2.6)**	**O2-O5/Mg25[a] (2.7)**	**N2-O5/Mg25[a] (3.5)** O6-O4/Mg22b (3.2)	**N2-O4/Mg25[a] (3.2)** N7-O4/Mg22b[g] (2.7)	G(20)
C(2)			N3-O2/Mg24c (3.4) O6-O4/Mg22a (2.8)	N7-O5/Mg22a (2.8)	G(19)
A(3)	N7-O2/Mg21 (2.7)		O2-O6/Mg24b (3.0)	O2-O2/Mg24c (2.8)	T(18)
G(4)	N2-O6/Mg24c (3.0) **N2-O4/Mg24a (3.5)** N2-O4/Mg24b (3.5) O6-O6/Mg21 (2.6)	N2-O2/Mg24a (3.1) **N2-O1/Mg24b (3.4)** N7-O3/Mg21 (2.7)	**O2-O4/Mg24a (2.4)**	**O2-O1/Mg24b (3.2)**	C(17)
G(5)	N2a-O5/Mg24a (2.9) N3b-O2/Mg24a (3.3)	N3a-O2/Mg24a (3.0)			C(16)

Contacts between all atoms in the bases and Mg^{2+} complexes are tabulated. Left/Right most columns indicate the "residue (residue number)/Chain ID". In the table, numbers in parenthesis display distance between the selected atoms in angstrom (Å). Small letters, a, b & c after each atom notates which alternate conformation the atom belongs to. First entry: N4-O2/Mg22a means the interaction between N4 in Residue 1 Cytosine (chain A) and 2nd coordinated water of Magnesium complex 22, conformation A. Bold letters indicate the common magnesium with contacts from both bases in the pair. Only top half of double stands is included. The bottom half is related by a 2-fold axis and contacts are identical. All surrounding magnesium complexes that appear within 3.5Å are included, including symmetry related ones. Super-scripted letters on the tabulated magnesium complexes indicates symmetry operations as listed below.

[a] *Magnesium complex on Special Position*
[b] *X-1/2, Y+1/2, Z*
[c] *-X, Y, -Z*
[d] *-X, Y+1, -Z*
[e] *-X+1/2, Y+1/2, -Z*
[f] *X+1/2, Y-1/2, Z*
[g] *X, Y, Z+1*

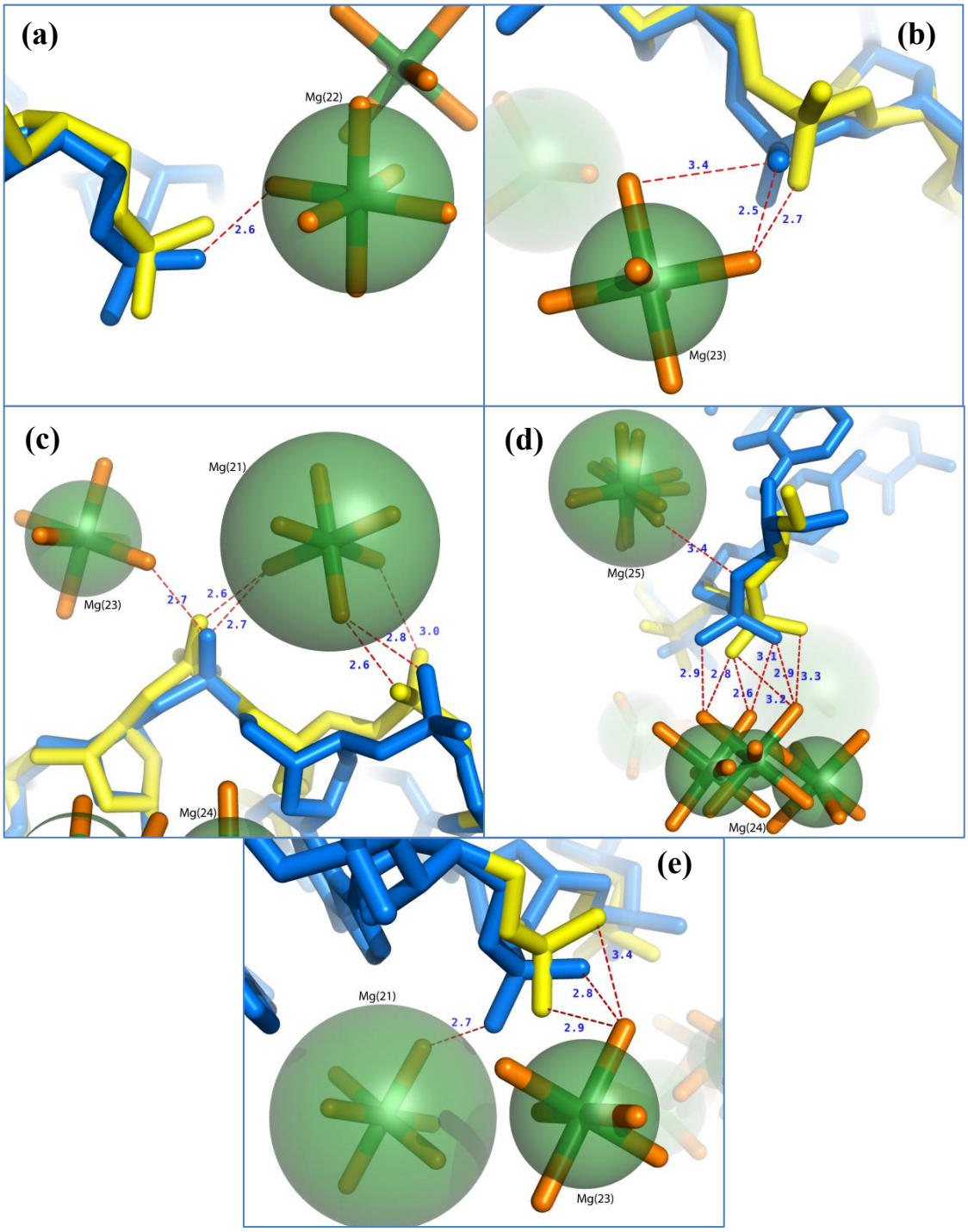

Figure 2.21: Correlation between alternating DNA states and neighboring Mg^{2+} ions. Figure (a) and (b) represents "Single cation Bound/Free mode" as described in the text. "Anchored Bound/Free mode" is shown in (c), (d) and (e), where Mg(21), Mg(25) and Mg(23) are anchoring the adjacent phosphate, respectively. Second Mg^{2+} at each location appear to correlate with the alternate states observed in the phosphate. Each Mg^{2+} ion is shown in stick, ***with the transparent-sphere indicating ion occupancy by its size****, relative to each Mg^{2+} in the same figure. Step shown in each figure is* ***(a)*** *G(4)pG(5),* ***(b)*** *C(7)pT(8),* ***(c)*** *G(5)pC(6) on the left and C(6)pC(7) on the right,* ***(d)*** *C(1)pC(2) and* ***(e)*** *C(2)pA(3) are shown. DNA coordinates are shown in (blue) alternate state "1" and (yellow) alternate state "2". Contacts between cation and phosphate backbone are shown in the range between 2.5-3.5Å. See plot 3.5, table 3.3 and 3.4 for the estimated occupancies and contacts details.*

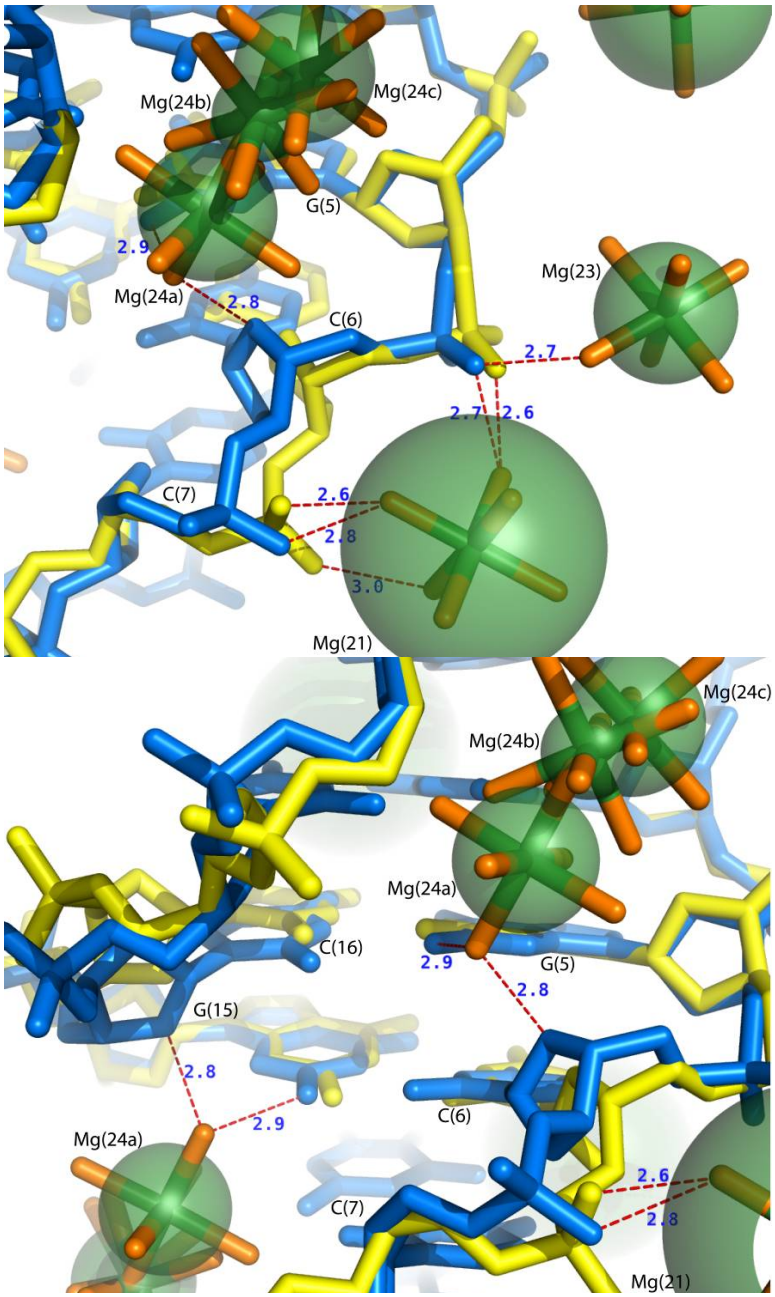

Figure 2.22: Correlation of disorders in the central G-tract and Mg^{2+}. (Top) Close up of contacts between hydrated magnesium ions and sugar/phosphate backbones. (Bottom) Close up of contacts with bases, viewing into minor groove at the same region. In both figures, DNA is colored blue for alternate state A, colored yellow for state B. Magnesium complexes are shown with ball and stick models, with metal centers colored green, and coordinated water molecules are colored orange. Contacts are shown in dotted line, with bond distances shown. Each magnesium ions are indicated with residue ID. Prime (') after the identifier indicates symmetry related complexes. Notice the state "A" specific contacts between N2 of G(5) and O4' of C(6) in the minor groove, while more number of contacts between state "B" phosphate group and neighboring magnesium complexes. Additionally, difference in occupancies between magnesium complexes may attribute to the resulting multiple alternating states of the DNA (Mg24a=40%, Mg21=90%) as described in the text.

Central GGCC step and Mg^{2+} complexes

The pronounced range of disorder observed at the central GGCC step also shows a similar correlation with the neighboring Mg^{2+} complexes (figure 2.22) as described earlier (Anchored Bound/Free mode). Here, Mg(21) acts as an anchoring Mg^{2+} creating hydrogen bonding interactions through the first shell water molecules with both phosphates of C(6) and C(7) regardless of the state. The second Mg^{2+}, minor-groove bound Mg(24a) flanks the backbone, making hydrogen bonding interaction through first shell water with O4' of the C(6) sugar when the DNA is in state A. Hence DNA state A (75% occupancy) is considered a bound state. The same Mg(24a) also makes a hydrogen bonding interaction with O6 of G(5) in state A. There also appears to be a third Mg^{2+} complex involved as Mg(23) associates with C(6) phosphate in state A only. Aside from the anchoring Mg(21), there is no hydrogen bonding interaction observed between free state DNA (state B) and Mg(24a) or Mg(23). Thus, the pronounced range of disorder displayed at the central GGCC step appears to correlate with the existence of partially occupied Mg^{2+} complexes, Mg(24) and Mg(23).

Mg^{2+} induced effects in local base parameters

The wide major groove found in B-DNA makes it possible for the hydrated ions to be shifted to one side of the groove, making contacts solely to one base within the pair (Tabel 3.5). The major groove Mg^{2+} complexes, 22 and 21 are found to form H-bond with both N7 and O6 of guanine residues at either strand, but not to their cytosine partners. These sites have been previously characterized as electrostatistically the most favorable loci (Pullman and Pullman 1981). Additionally, Mg(21) also makes H-bond contact with N7 of adenine residue at the A(3)pG(4) step. Hence, the positive roll angle

observed at the terminal CpC step and the A(3)pG(4) appear to link with the neighboring Mg(22) and Mg(21) respectively, as each Mg^{2+} complex makes intra-strand contact with both guanine residues or guanine and adenine in latter case. The effect of the major-groove bound Mg^{2+} also appears to extend to the variations seen in the base-base parameters within these steps. As seen in figure 2.12, high propeller twist is observed at the terminal CG pair, and wide opening A(3)/T(8) pair is also observed. However, such local effects may need to be viewed as a coupled event between these major-groove bound cations and also the relative positions of the minor-groove Mg^{2+} complexes. Firstly, these major-groove bound Mg^{2+} complexes make contacts to the only one base of each pair (guanine or adenine) in the major groove, not to their pairing residues. Secondly, the variation in base-base parameters, such as propeller twist and openings appear to be determined by the relative positions of minor-groove bound Mg^{2+} complexes, which is displayed by the contacts with Mg(25) for the terminal cytosine and Mg(24) for the T(8). For example, high positive roll of the terminal CG pair may be the result of intra-strand major-groove contacts between N7/O6 of both guanine residues and Mg(22), coupled with minor-groove contact between O2/C(1) and Mg(25).

Hydration

Given the atomic resolution of the current data, majority of the water molecules have been identified. Alternate networks of 32 fully occupied and 61 partially occupied water molecules (53.2 total occupancies) per asymmetric unit nearly fill the solvent channel. In the minor groove, the center is sparsely hydrated as noted originally (Heinemann and Alings 1989) (see figure 2.23). A water molecule HOH (228) is positioned between base planes of C(6) and C(16), slightly off-set from the

crystallographic special position. Hence its symmetry equivalent molecule is 0.7Å apart. At either position, HOH(228) makes inter-strand contacts with O2 of C(6) and C(16). Aside this, partially occupied water molecules (259 and 246) are present at the center of the groove, and 25% occupied HOH(246) makes the hydrogen bonding contact with O2P of the T(8) when DNA is in alternative state B. As HOH(246) and HOH(259) are positioned at 2.4Å apart, these partially occupied water molecules cannot present at the same time. This short pseudo-string of hydration is disrupted by 2 flanking Mg^{2+} complex 24 (each with 3 alternative states) in the minor groove, replacing the hydration of bases with their first shell water molecules (see table 2.5). Hydration of the minor groove reappears toward the terminal base pairs, subsequently divided into two columns of waters interacting with either strand of the helix. The hydrogen bonding interactions with these two columns of water molecules involve with N3 of terminal G/C pair, O2 of C(2), and N3 of both A(3) and G(4). Then, again the minor groove hydration is interrupted by the Mg^{2+} complex 25. Additionally, some of the water molecules also make hydrogen bonding interaction with O4', similar to central double-spine observed previously reported (Heinemann and Hahn 1992; Baikalov, Grzeskowiak et al. 1993).

The pattern of ordered hydration in the major groove appears to be in agreement with consensus hydration sites of B-DNA (Schneider and Berman 1995). A water molecule lies within a hydrogen bond distance to the polar group of each, including N4 of cytosine, O4 of thymine, N7 and O6 of guanine, and N7 and N6 of adenine. However, first shell water molecules of neighboring Mg^{2+} complexes often replace these consensus hydration sites for guanine residues. At G(4) and G(9), both N7 and O6 are hydrated by 2 first shell water molecules of neighboring Mg^{2+} (see table 3.5). Because of the absence of

Mg^{2+} complex at the center of the duplex, G(5)/G(15) are the only guanine residues showing no Mg^{2+} interaction in the major groove.

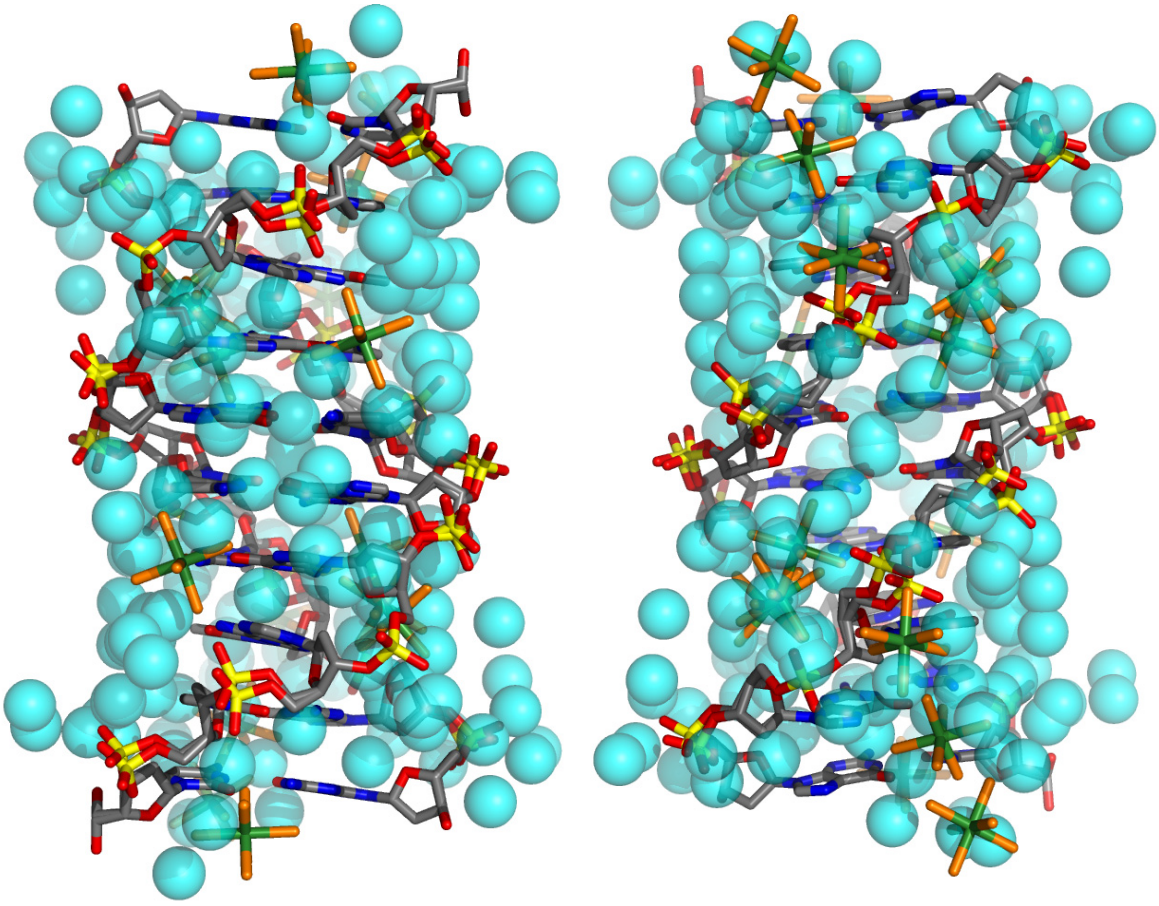

Figure 2.23: Hydration of [d(CCAGGCCTGG)]$_2$. Viewing into (Left) the major groove (Right) the minor groove. Water molecules are shown in transparent sphere, colored cyan.. DNA is shown with stick-model, colored in CPK coloring scheme. Hexaaqua Mg ions are shown in stick-model, colored green at its metal center and orange for the coordinated water molecules. Hydration patterns are largely disrupted by the presence of hexaaqua Mg2+ complexes in both major and minor grooves. Notice the sparse hydration in both grooves. See text.

Possibility of localized monovalent cations

Accounting for charge

The charge imbalance between DNA and cations indicates that some of the counter-ions are missing from the model. In the current structure, a total of ten sites are partially occupied by hydrated magnesium complexes (per duplex). These ten partially occupied Mg^{2+} ions translate into positive charge of +13.2 *Coulombs*. The total positive charge is determined by using the refined occupancies of the Mg^{2+} ions and +2 *Coulombs* per Mg^{2+} metal center. The net charge of the DNA plus Mg^{2+} ions is Δq = -4.8. There are -18 *Coulombs* arising from eighteen phosphate groups of the DNA duplex. Therefore, there is charge imbalance between identified counter-ions and DNA. Na^+ and Mg^{2+} are the only cations present in the crystallization solution.

Na^+ ion

It is most likely that the unobserved cations are Na^+ ions, because Mg^{2+} ion is readily identifiable by coordination geometry whereas Na^+ is not. Hydrated form of Na^+ is known to take somewhat irregular coordination geometry (typically still in hexa-coordination) when compared to a nearly perfect octahedron of hydrated Mg^{2+} (Phillips, Dauter et al. 1997). The bond distance between its atom center to coordinated water is also longer at approximately 2.4-2.6Å (compared to ~2.1Å of hydrated Mg^{2+}). Hydrated form of Na^+ has been previously reported that its metal center is also capable of direct coordination to O6 of guanine base, replacing one of its coordinated water (Komeda, Moulaei et al. 2006). In the current structure, no cluster of electron densities that fits the criteria above for a hydrated form of Na^+ is recognized.

In the case of partially-dehydrated Na^+ ions, it is nearly indistinguishable in the electron density maps. It is largely due to the fact that Na^+ has the same number of electrons as water (10 electrons) and the shape of the electron density is the same (spherical). Furthermore, the difference in ionic/molecular radii between Na^+ and water (0.95Å vs. 1.4Å) generally does not contribute enough to discern these two. Because the effective dispersion of the electrons about the atom center varies with thermal fluctuations, positional disorder and data quality, these effects obscure the difference in effective radii (Williams 2005).

In the final refined model, again, no sodium ion was identified through coordination geometry. Therefore, due to the limitations imposed by current method as mentioned above, all the independent solvent peaks other than those for magnesium complexes are arbitrary assigned as water molecules. In later chapter, the use of monovalent cations with strong scattering characteristics, namely Thallium (I) and Rubidium Ions is described, with the aim to locate localized monovalent cations within the same sequence of DNA.

Conclusion

The ultra-high resolution structure of $[d(CCAGGCCTGG)]_2$ reveals greater details on structural flexibility within such helices that are undetected in its lower resolution structure. The frequently observed alternate states in the current structure may also provide insights to high thermal factors among sugar-phosphate backbone typically observed in lower resolution structures. On the other hand, it also raises a concern among recently published ultra-high resolution structures as well. Static disorders in sugar-phosphate backbone have been reported recently in isomorphous *B*-DNA structures of

$[d(CCAGTACTGG)]_2$, $[d(CCAGCGCTGG)]_2$ and $[d(CCAACGTTGG)]_2$, all with data that extends beyond 1Å (Chiu and Dickerson 2000; Kielkopf, Ding et al. 2000). While each work displays low average isotropic thermal factors (*B*-value) through DNA coordinates, few phosphate groups still show unusually high value between their alternate states. Coincidently, such deviation is seen at phosphate of C(7), where the pronounced structural variation is observed in the current structure. The high *B*-value regions within these structures also correspond to the abnormal shape of thermal ellipsoids reported as well. *B*-value directly correlates with the thermal vibration or static disorders as it is a measure of displacement of an atom about its average position. With more defined electron densities provided in atomic resolution data, it is possible to assign alternate model effectively. Target atomic displacement is thus reduced, and resulting *B*-values for both alternate models must correspond to such treatment. Some alternate model assigned in these aforementioned high resolution structures contradict with resulting *B*-values, as shown in the large deviations from the rest of the model. This may thus indicate the inaccurate alternate models that were assigned.

In practice, the assignment of alternate model should be reasonably explained by the existence of both $2F_o$-F_c and F_o-F_c peaks. At reduced occupancy, some of the low-electron atoms, such as carbon in sugar may display weak $2F_o$-F_c peaks. However, the connecting atoms with higher number of electrons, such as oxygen, should provide the traceability of the backbone. Conversely, even at atomic resolution, experimental evidence for placement of hydrogen atom is not expected other than F_o-F_c peaks (Kielkopf, Ding et al. 2000; Wlodawer, Minor et al. 2008). In a given region where an

alternate model is thought, crystallographer should thus carefully analyze in terms of occupancy, atom type and its relative appearance in $2F_o$-F_c and F_o-F_c peaks.

Providing high atomic resolution of data, the crystallographic *R*-factor is effectively a measure of model error. However, resulting *R*-factor may not be sufficiently indicative of misinterpreted local region. In fact, *R*-factor may itself display reduced value as occupancy is cut at original over-assigned region, or as its alternate model partially overlaps with the region in necessity of a new model assignment. Hence, careful analysis of individual isotropic *B*-factor and anisotropic displacement (thermal ellipsoid) should be practiced, as abrupt deviation may well indicate the erroneous interpretation.

Database accession number

The structure of $[d(CCAGGCCTGG)]_2$ has been assigned PDB ID of 3GGB.

CHAPTER 3

LOCATING MONOVALENT CATIONS IN ONE TURN OF G/C RICH B-DNA

Abstract

In this chapter, we describe ultra-high resolution x-ray structures of a duplex of CCAGGCCTGG obtained from DNA crystals grown in the presence of thallium (I) or rubidium ion. Twelve Tl^+ sites and four Rb^+ sites are identified in the respective $[d(CCAGGCCTGG)]_2/Tl^+$ and $[d(CCAGGCCTGG)]_2/Rb^+$ crystals. Positions of Rb^+ ions in the refined model are consistent with Tl^+ positions identified in this work. Both Tl^+ and Rb^+ positions are predominantly within the grooves, and are not in direct proximity to phosphate groups. Majority of Tl^+/Rb^+ positions are in the major groove of CCAGGCCTGG, coordinated by N7 and/or O2 of adjacent guanine residue. Some of the major groove bound Tl^+/Rb^+ are also coordinated by the second N7 of preceding purine from the same strand, which was not seen in the previous study with CGCGAATTCGCG $/Tl^+$. The minor groove of the duplex of CCAGGCCTGG contains two Tl^+ sites at the TpG step of each strand, coordinated by O2 of a thymine and O4' of a following guanine residue. Cross-strand chelations by the second O2 of thymine and O4' from the opposing strand seen in the A-tract minor groove of CGCGAATTCGCG was not observed. Tl^+/Rb^+ sites determined in CCAGGCCTGG hence appear to hold a similar mode of coordination observed previously in CGCGAATTCGCG. Differences observed between Tl^+/Rb^+ positions in this study and the previously determined Tl^+ positions in the duplex

of CGCGAATTCGCG may arise from the difference in DNA conformation that is sequence dependent.

Introduction

In recent years, compelling evidence has emerged for that monovalent cations can localize around *B*-DNA in geometrically regular and sequence-dependent manner. Such evidence has come from x-ray crystallography (Shui, McFail-Isom et al. 1998; Tereshko, Minasov et al. 1999; Woods, McFail-Isom et al. 2000; Howerton, Sines et al. 2001; Tereshko, Wilds et al. 2001; Howerton, Nagpal et al. 2003), solution NMR (Hud, Sklenar et al. 1999; Denisov and Halle 2000; Cesare Marincola, Denisov et al. 2004) and molecular dynamics (MD) simulation (Young, Jayaram et al. 1997; Hamelberg, McFail-Isom et al. 2000; McConnell and Beveridge 2000; Hamelberg, Williams et al. 2002; Mocci and Saba 2003). The possible correlations between localized monovalent cations and sequence-dependent variation of DNA conformations have been discussed this past year (Rouzina and Bloomfield 1998; McFail-Isom, Sines et al. 1999; Williams and Maher 2000; Hud and Polak 2001; Egli 2002; Hud and Plavec 2003)

We have previously solved 0.96Å resolution structure of $[d(CCAGGCCTGG)]_2$ through x-ray crystallography (see chapter 2). Despite the quality of data and careful analysis, charge imbalance was apparent between DNA and identified cations indicating some of the counter-ions are missing from the model. We concluded that it is most likely that the unobserved cations are Na^+ ions, because Mg^{2+} ion is readily identifiable by coordination geometry whereas Na^+ is not.

Here, ultra-high resolution structure of $[d(CCAGGCCTGG)]_2$ derivatives are obtained from crystals grown in the presence of thallium (Tl^+) or rubidium (Rb^+) along

with Mg^{2+}. The advantage of the use of Tl^+ and Rb^+ in x-ray crystallographic studies is that they are directly observable because of their strong scattering of x-rays and their anomalous scattering. These ions are hence identifiable without resorting to analysis of subtle difference in coordination geometry. Tl^+ and Rb^+ have been used effectively to probe cation binding site in macromolecular crystallography. Tl^+ was used to locate monovalent cation site in the crystal of a catalytic RNA (Basu, Rambo et al. 1998), $[d(CGCGAATTCGCG)]_2$ or its variant (Howerton, Sines et al. 2001; Moulaei, Maehigashi et al. 2005), ribosomal RNA fragment (Correll, Wool et al. 1999; Conn, Gittis et al. 2002), G-quadruplex (Gill, Strobel et al. 2006) and a complex of N-terminal domain of P22 c2 repressor protein and its operator sequence (J.D. Watkins and L.D. Williams, unpublished). Rb^+ was used in the same manner in macromolecular crystallography including for a membrane bound K^+ channel (Doyle, Morais Cabral et al. 1998), $[d(CGCGAATTCGCG)]_2$ (Tereshko, Minasov et al. 1999), $[d(GCGTATACGC)]_2$ (Tereshko, Wilds et al. 2001), a ribosomal subunit (Klein, Moore et al. 2004) and a membrane bound Na^+/K^+ pump (Morth, Pedersen et al. 2007). Our observations of Tl^+/Rb^+ interactions with $[d(GCGTATACGC)]_2$ are very similar to previous studies, which is Tl^+ and Rb^+ ions make contacts primarily with uncharged macromolecular oxygen and nitrogen atoms and water. Majority of Tl^+/Rb^+ positions are in the major groove of the duplex of CCAGGCCTGG; however the minor groove also contains two Tl^+ sites (figure 3.1-3.2).

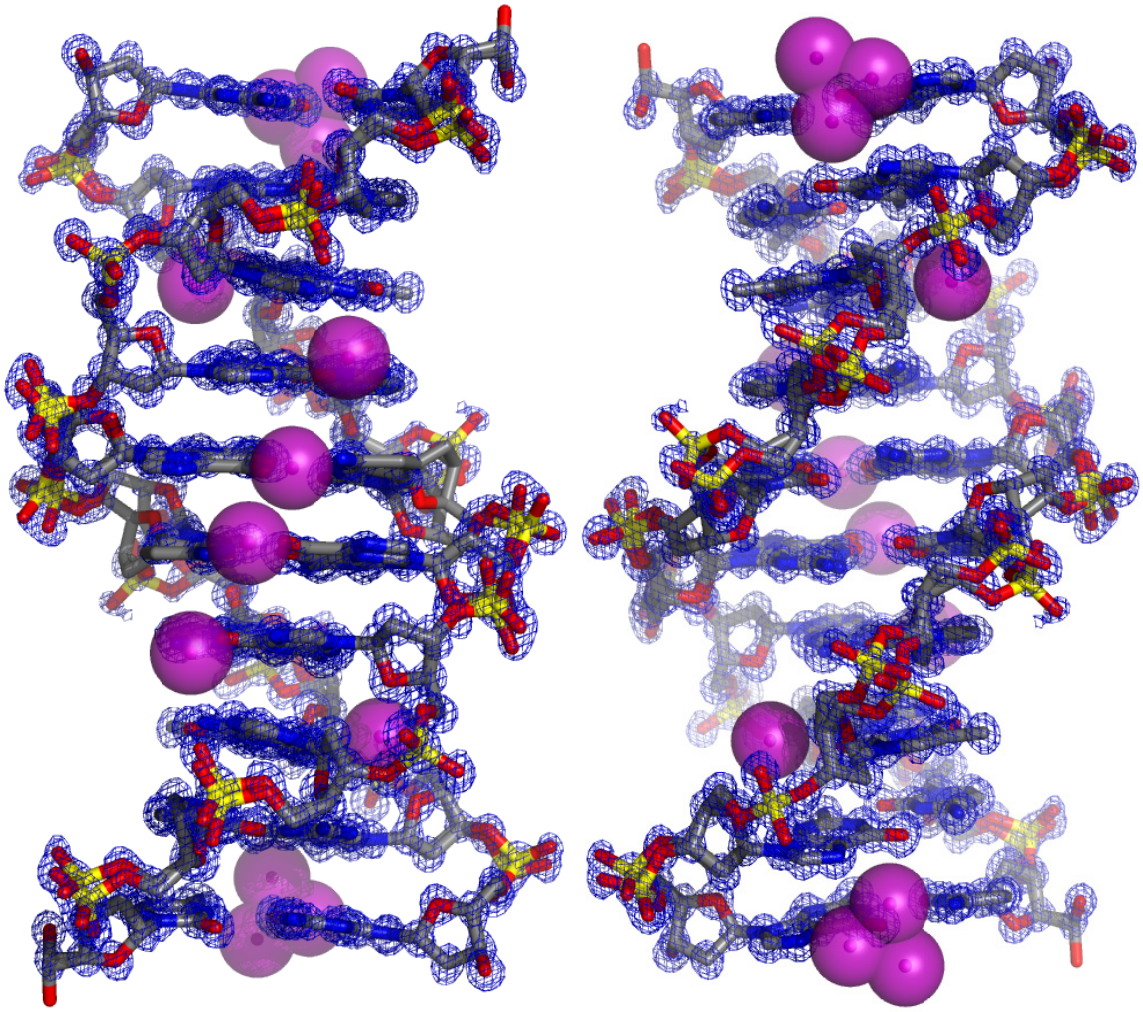

Figure 3.1: Overall view of the 0.98Å crystal structure of [d(CCAGGCCTGG)]$_2$/Tl$^+$. Viewing into major groove (left) and minor groove (right). DNA strands are shown in stick models, colored with CPK coloring scheme. Sum electron density ($2F_o$-F_c) is shown in blue net around DNA, contoured at 1.0σ. Thallium (I) ions are shown in purple spheres. Backbone of the helix show high static ordered region, similar to that of Na$^+$ form of the same oligomer (see chapter 2). Hexaaqua magnesium complexes and water molecules are omitted for clarity. Unless otherwise noted, all figures in this chapter are rendered with Pymol v1.1 (DeLano).

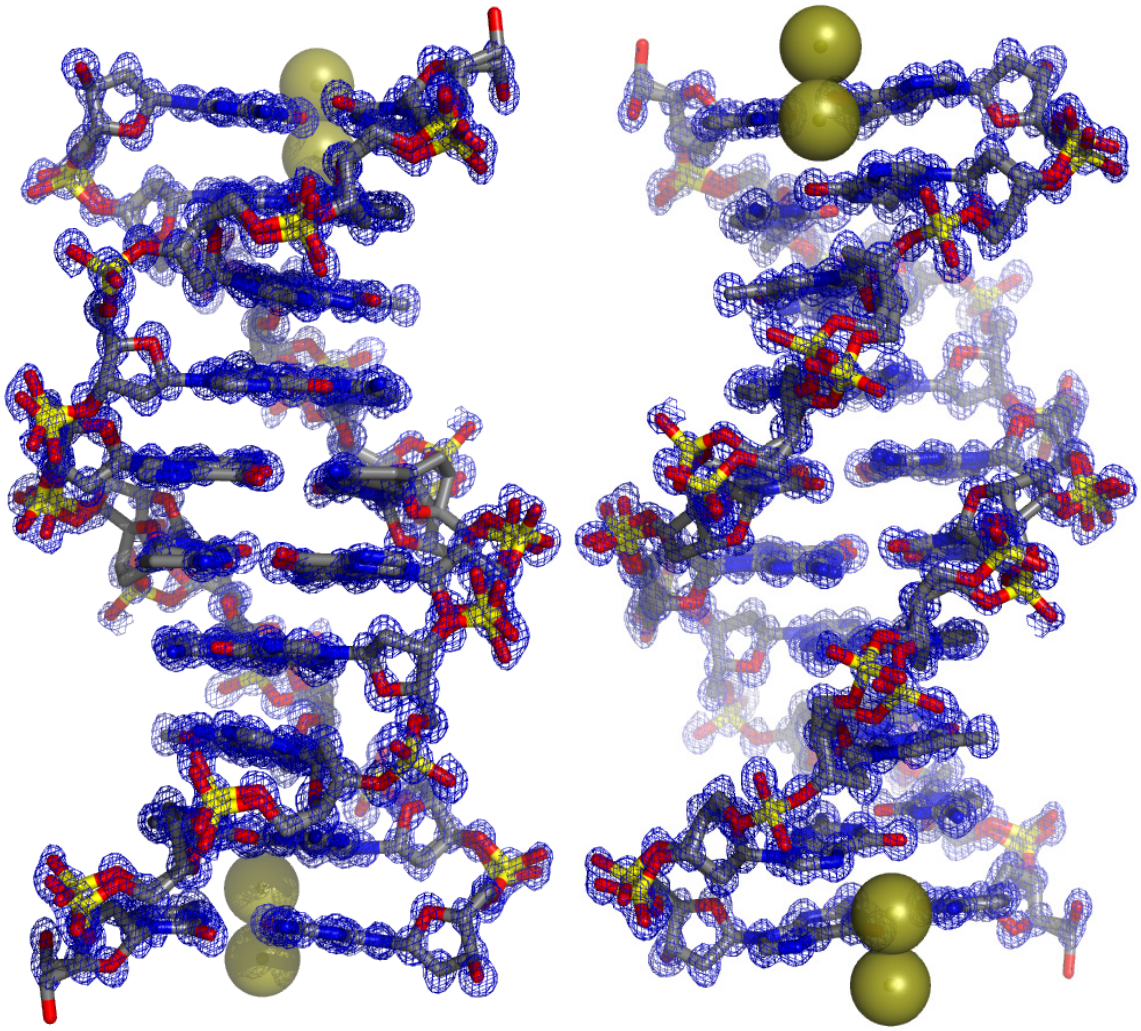

Figure 3.2: Overall view of the 0.87Å crystal structure of [d(CCAGGCCTGG)]$_2$/Rb$^+$. Viewing into major groove (left) and minor groove (right). DNA strands are shown in stick models, colored with CPK coloring scheme. Sum electron density ($2F_o$-F_c) is shown in blue net around DNA, contoured at 1.0σ. Rubidium (I) ions are shown in olive spheres. Hexaaqua magnesium complexes and water molecules are omitted for clarity. Unless otherwise noted, all figures in this chapter are rendered with Pymol v1.1 (DeLano).

Materials and Methods

Crystallization and Data Collection

Reverse-phase HPLC purified d[CCAGGCCTGG] (Integrated DNA Technology) was purchased, and annealed slowly from 90°C in a water bath upon reconstitution to the desired concentration in nano-pure water. Crystals of Tl^+ and Rb^+ salts of $[d(CCAGGCCTGG)]_2$ were grown by hanging drop vapor diffusion method. In both cases, sodium cacodylate buffer found in crystallization solution (used for Na^+ salt of decamer) was replaced with pH adjusted Tl^+ or Rb^+ acetate solution to eliminate the competing monovalent cations. Briefly, thallium (I) acetate was titrated with cacodylic acid, and rubidium acetate was titrated with rubidium hydroxide, both to final pH of 6.5.

Tl^+ salts of $[d(CCAGGCCTGG)]_2$

Crystallization drop initially contained 0.23mM (single strand base) of d(CCAGGCCTGG), 12.5mM magnesium acetate, 15mM thallium (I) acetate (pH6.5) and 17.5% (v/v) 2-Methyl-2,4-pentanediol (MPD). The crystallization drop was equilibrated against the reservoir solution containing 35% (v/v) MPD at 4°C.Plate-like crystals with the dimension approximately 0.2 x 0.1 x 0.05mm^3 appeared, typically in 5 -6 days at 4°C. The crystals displayed high birefringence character and edges appeared sharp.

Crystals were looped with 0.1-0.2mm 20micron nylon cryoloop (Hampton Research, California, USA), followed by flash-freeze in the liquid nitrogen and kept frozen until the data collection. x-ray diffraction data used for the refinement was collected at beamline 22-ID in the SER-CAT facilities at Advanced Photon Source (Argonne, IL, USA) for a total of 360° of data with 1° oscillation using MAR 300 CCD detector (Marresearch GmbH, Germany) using 0.975Å radiation (Table 3.1). The crystals

were cryogenically cooled at -160°C during data collection. These crystals diffracted strongly, with frequent overload to the detector even at mechanically applicable minimum exposure time (1 second) at each oscillation, hence a thin metal foil was applied to the path of incoming beam to minimize the damage to the detector. A total of 94406 reflections through 360 frames were collected, indexed and reduced to 13522 unique reflections at the maximum resolution of 0.92. The complex of Tl^+ - $[d(CCAGGCCTT)]_2$ crystallizes in the monoclinic space group C2, with unit cell parameters of a=32.28Å, b=25.20Å, c=34.10Å, α=γ=90° and β=116.10°. The table 3.1 summarizes data collection and refinement statistics.

Rb^+ salts of $[d(CCAGGCCTGG)]_2$

Crystallization drop initially contained 0.23mM (single strand base) of d(CCAGGCCTGG), 12.5mM magnesium acetate, 20mM rubidium acetate (pH6.5) and 22.5% (v/v) 2-Methyl-2,4-pentanediol (MPD). The crystallization drop was equilibrated against the reservoir solution containing 45% (v/v) MPD at 4°C. Plate-like crystals with the dimension approximately 0.2 x 0.1 x 0.05mm^3 appeared, typically in 5 -6 days at 4°C. The crystals displayed high birefringence character and edges appeared sharp.

Data collection procedure is identical to Tl^+ salts as described above, other than 0.82Å radiation for Rb^+ salts (Table 4.1).A total of 119366, reflections through 360 frames were collected, indexed and reduced to 19281 unique reflections at the maximum resolution of 0.78. The complex of Rb^+-$[d(CCAGGCCTT)]_2$ crystallizes in the monoclinic space group C2, with unit cell parameters of a=32.14Å, b=25.16Å, c=34.11Å, α=γ=90° and β=116.24°. The table 4.1 summarizes data collection and refinement statistics.

All data (both Tl^+ and Rb^+ salts) were processed and scaled using HKL2000 package (Otwinowski and Minor 1997) with anomalous flag turned on. In this option, anomalous pairs are considered equivalent when calculating scale B factors when computing statistics, however Bijovets pair (I_+ and I_-) are kept separate in the output file.

Refinement

The phase determination was carried out by a molecular replacement method with the DNA coordinates from $d(CCAGGCCTGG)_2$ at 1.6Å [NDB entry BDJ017 (Heinemann and Alings 1989)] using program *CNS version 1.1* (Brunger, Adams et al. 1998) along with parameters of Berman and co-workers (Berman, Zardecki et al. 1998). The improved starting model that consists of only DNA from the program *CNS 1.1* through a several round of simulated annealing and refinement was directly fed to the program *REFMAC5* (Murshudov, Vagin et al. 1997). A care was taken in terms of preserving the flagged intensities for Free-*R* test between these two programs. The original intensity file (.sca) was converted to MTZ using the program sca2mtz first, and then this MTZ output was processed by the program Mtz2various to be used in *CNS*. Both utilities are available in *CCP4 program suit* (Bailey 1994).The same MTZ file was used for the refinement process by the program *REFMAC5*. Hydrogen atoms have been added in the riding positions during the refinement process (with no output to the final PDB). Water molecules were added to peaks of corresponding $2F_o$-F_c and F_o-F_c peaks visualized in the program *Coot* (Emsley and Cowtan 2004). Some water molecules were replaced with partially occupied Tl^+ or Rb^+ ions only if both anomalous (F_+-F_-) difference and isomorphous (F_{Tl+} or F_{Rb+} -F_{Na+}) difference peaks are observed at the site. As discussed earlier, the program *SHELXL* (Sheldrick 1997; Sheldrick 2008) was utilized in

initial assignment of occupancies to the regions where alternative states/partial occupancies were observed. All atoms in the model were refined anisotropically. Refinement process was continued until the convergence of refinement statistics was reached (Table 3.1), while R-free and thermal ellipsoids were monitored to avoid over refinement. Thermal ellipsoids (figure 3.3a, 3.3b) were computed with *ORTEP-3 for windows* (Farrugia 1997).

Multi-state model of DNA and Mg^{2+}

Procedure for identification and treatment of multi-state model of DNA and Mg^{2+} complexes are identical to the previously solved Na^{+} salt of the decamer (see chapter 2). The program *Coot* (Emsley and Cowtan 2004) was utilized to the region where manual build of the additional model was necessary.

Anomalous ($|F_{+}-F_{-}|$) difference maps

Determination of sites occupied by monovalent cations using anomalous scattering is similar to Single-wavelength Anomalous Diffraction (SAD) experiment. However, the procedure is somewhat reversed in our cases as the phase information is already known through molecular replacement, yet locations of heavy atom sites are unknown. The $|F_{+}-F_{-}|$ difference Fouriers were calculated for $[d(CCAGGCCTGG)]_2/Tl^{+}$ and $[d(CCAGGCCTGG)]_2/Rb^{+}$ using program *CNS version 1.1* (Brunger, Adams et al. 1998). Phases for the anomalous difference maps were calculated using entire respective models with all solvent models removed in both Tl^{+} and Rb^{+} cases.

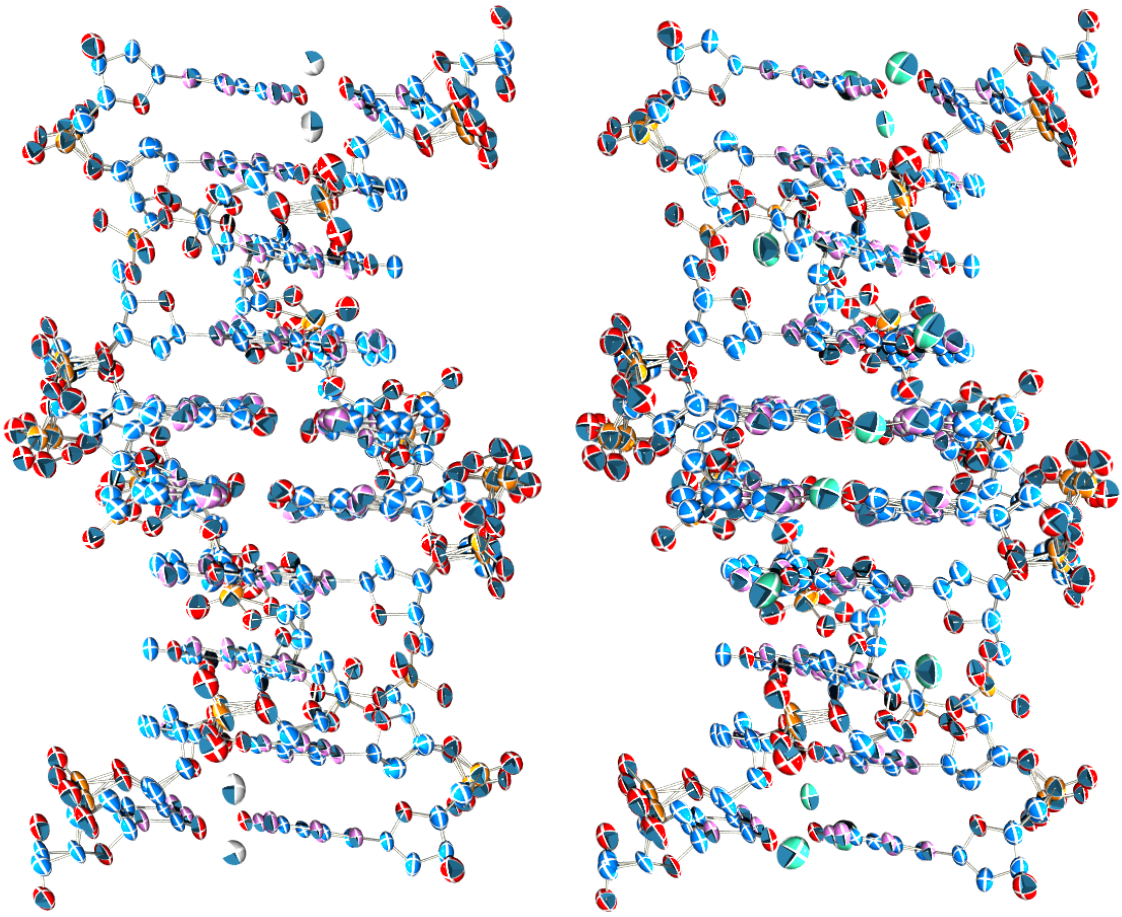

Figure 3.3a: Comparison of thermal ellipsoids between two ultra-high resolution crystal structures of duplex of B-DNA, CCAGGCCTGG. Both viewing into the major groove. Two derivative forms are shown, Tl^+ (Left) salt at 0.98Å resolution, and Rb^+ (Right) salt at 0.87Å resolution Both plots are at 50% probability. Tl^+ ions are shown in light-green on left. Rb^+ ions are shown in light-gray. Thermal ellipsoids shown in two structures are comparable, without any specific area that deviates greatly. For reference, the average isotropic b-values (atomic displacement) are also comparable. 7.86Å^2 and 7.76Å^2 for Tl^+ and Rb^+ form, respectively. See chapter 1 for the discussion on this matter. The plots were calculated with Ortep-3 for windows (Farrugia, 1997) and rendered with Povray 3.6. Hexaaqua magnesium complexes and water molecules are omitted for clarity.

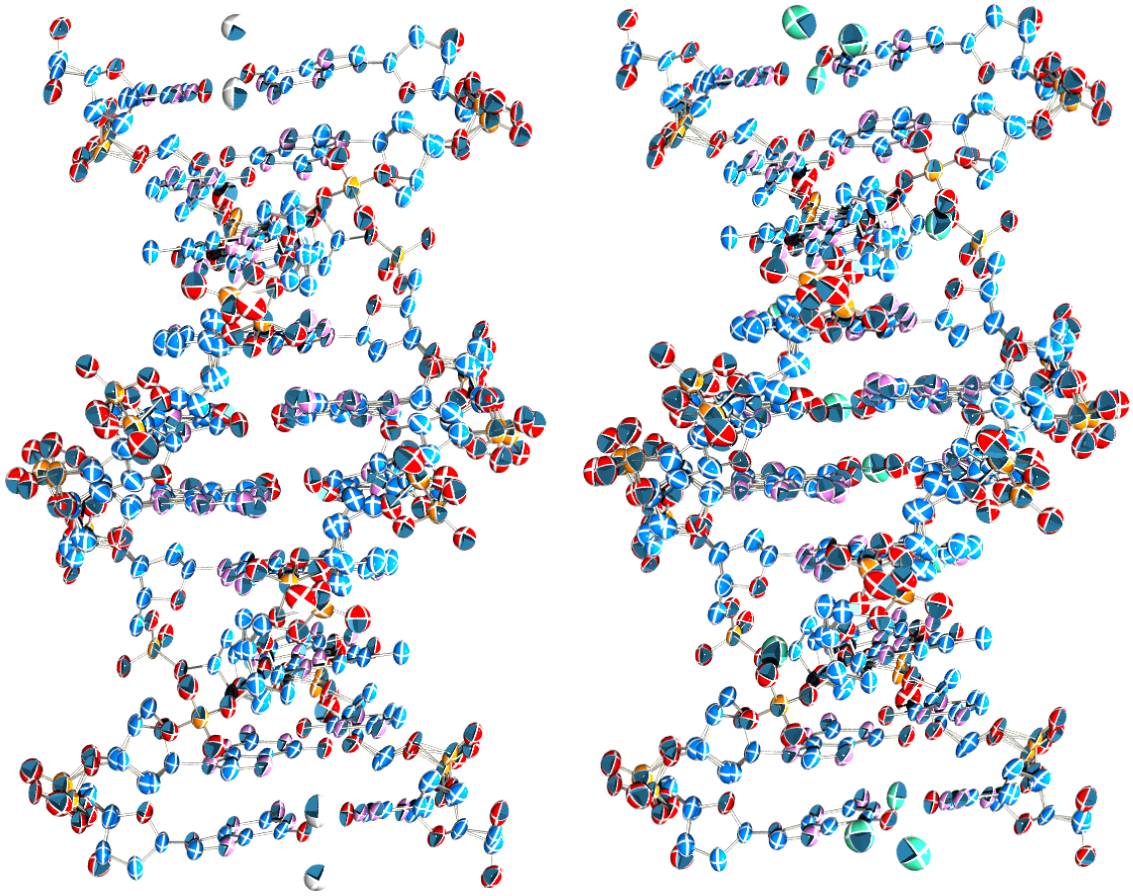

Figure 3.3b: Comparison of thermal ellipsoids between two ultra-high resolution crystal structures of duplex of B-DNA, CCAGGCCTGG. Both viewing into the minor groove. Two derivative forms are shown, Tl^+ (Left) salt at 0.98Å resolution and Rb^+ (Right) salt at 0.87Å resolution. Both plots are at 50% probability. Tl^+ ions are shown in light-green on left. Rb^+ ions are shown in light-gray. The plots were calculated with Ortep-3 for windows (Farrugia, 1997) and rendered with Povray 3.6. Hexaaqua magnesium complexes and water molecules are omitted for clarity.

Table 3.1: Crystallographic and Refinement Statistics

	$[d(CCAGGCCTGG)]_2/Tl^+$	$[d(CCAGGCCTGG)]_2/Rb^+$
Data Collection		
Source wavelength (Å)	0.97582	0.81528
Space group	C2	C2
Unit cell	a=32.277Å	a=32.135Å
	b=25.204Å	b=25.156Å
	c=34.095Å	c=34.095Å
	α=γ=90° β=116.10°	α=γ=90° β=116.24°
Resolution range (Å)	15.3-0.98	30.60-0.87
Number of reflections (all)	94406	119366
Number of unique reflections	13011	18590
Completeness (%)[a]	90.05 (52.88)	90.55 (50.26)
Average I/σ (I)	61.0	23.0
R_{merge} (%)[b]	14.3	7.4
Refinement statistics		
DNA (asymmetric unit)	d(CCAGGCCTGG)	d(CCAGGCCTGG)
Number of DNA atoms	287	287
Number of water molecules[c]	76	91
Number of heavy atoms (Tl^+ or Rb^+)[c]	6	2
Number of hexaaqua Mg ions	5	5
R_{work} (%)[d]	11.8	11.9
R_{free} (%)[e]	14.4 (992 reflections)	13.4 (1047 reflections)
RMS deviation of bonds from ideal	0.010	0.010
RMS deviation of angles from ideal	1.892	1.833
Average isotropic B value ($Å^2$)	7.856	7.764
PDB ID code	3GGI	3GGK

[a] *The values in parentheses refers to the highest resolution shells*
[b] $R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma I$, *where I = observed intensities and* $\langle I \rangle$ *= mean intensity obtained from multiple observations of symmetry-related reflections and rejections.*
[c] *Includes partially occupied species as well as molecules on special position. Disordered species with same residue number are counted as single molecules.*
[d] $R_{work} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$, *where Fo and Fc are the observed and calculated structure factors, respectively. Reflections flagged for Free R test (7.8%) are excluded from the calculation. R factor for all reflections is determined to be 11.97%.*
[e] R_{free} *= defined by Brünger (Brunger 1992)*

Isomorphous (F_{Tl+} or F_{Rb+}-F_{Na+}) difference maps

Additionally, isomorphous (F_{Tl+} or F_{Rb+} - F_{Na+}) difference Fouriers were utilized to confirm the heavy atom sites determined through anomalous difference maps with the diffraction data from high resolution $[d(CCAGGCCTGG)]_2/Na^+$ structure (PDB ID, 3GGB) (see chapter 3). The F_{Tl+} or F_{Rb+} - F_{Na+} difference Fouriers were calculated using program *CNS version 1.1*(Brunger, Adams et al. 1998) with the phase information calculated using respective models with all solvent models removed. Each set of data (Na^+/Tl^+ and Na^+/Rb^+) was merged (*CNS: merge.inp*) and each derivative data was scaled to the Na+ data (*CNS: scale.inp*) prior to map calculations. The observed R_{iso} (Crick and Magdoff 1956) upon merging each set of data are 0.203 (Na^+-Tl^+) and 0.137 (Na^+-Rb^+) averaged over all shells.

Providing low R_{iso} observed between each set of data, DNA coordinates from $[d(CCAGGCCTGG)]_2/Na^+$ crystal is highly isomorphous to DNA found in both $[d(CCAGGCCTGG)]_2/Tl^+$ and $[d(CCAGGCCTGG)]_2/Rb^+$ crystals. With similar dimensions observed in unit cell parameters of max Δ=0.4% (Tl^+/Na^+) and max Δ=0.1% (Rb^+/Na^+) and RMS deviation (all DNA atoms) of 0.04Å (Tl^+/Na^+) and 0.05Å (Rb^+/Na^+) indicate that DNA coordinates are nearly identical in these structures.

Results

The final $[d(CCAGGCCTGG)]_2/Tl^+$ model contains six partially occupied monovalent cation sites per single strand (asymmetric unit). The final $[d(CCAGGCCTGG)]_2/Rb^+$ model contains two partially occupied monovalent cation sites per single strand (table 3.2). Both models contain five disordered and/or partially occupied hexaaqua Mg^{2+} ions at the sites nearly identical to the previously solved $[d(CCAGGCCTGG)]_2/Na^+$. During refinement, all corresponding $2F_o$-F_c and F_o-F_c peaks other than for DNA coordinates and Mg^{2+} complexes were initially assigned as water molecules. This is to avoid bias of the final model toward Tl^+ or Rb^+ ions over water molecules. Conversions of some of the water molecules to the partially occupied monovalent cations were only performed if the site satisfies all criteria described below. (a) Significant residual F_o-F_c is peak is observed on top of the $2F_o$-F_c peak of the refined water molecule. This indicates the necessity for increased scattering by possibly by heavy atoms. (b) The tentative heavy atom site in the refined structure is also signified by the presence of anomalous F_+-F_- peak. The anomalous difference Fourier was calculated with phases from the respective model with solvent removed to avoid bias by positions of heavy atoms in the refined structure. (c) The intensity of the anomalous F_+-F_- peak must be greater than the residual noise, defined as the peak that appears irrespective of any atomic site. Hence anomalous peak height must be greater than 4.5σ in $[d(CCAGGCCTGG)]_2/Tl^+$ and 4.2σ in $[d(CCAGGCCTGG)]_2/Rb^+$. (d) The appearance of the anomalous F_+-F_- peak is consistent with the isomorphous F_{Tl+}-F_{Na+} or F_{Rb+}-F_{Na+} peak that overlaps each other. The isomorphous difference Fourier was calculated with phases from the respective model with solvent removed and diffraction data from high resolution

Table 3.2:Tl^+/Rb^+ difference peak intensities and location in $[d(CCAGCGCTGG)]_2$

ID	Occu. (%)	*iso-b*[a] ($Å^2$)	$\|F_+-F_-\|$ peak[b]	Δd_{peaks}[c] (Å)	$F_{Tl}/F_{Rb}-F_{Na}$	Δd_{peaks}[e] (Å)	F_o-F_c peak[f]	Location
Tl^+ 101	24	7.29	42.0	0.10	50.4	0.17	29.3	Terminal GG major groove
Tl^+ 102	11	14.08	11.5	0.21	13.3	0.19	10.6	G-tract major groove
Tl^+ 103	16	11.42	14.1	0.17	17.3	0.16	11.7	G-tract major groove
Tl^+ 104	6	11.73	6.8	0.30	6.4	0.18	4.5	Terminal GG major groove
Tl^+ 105[SP]	6[SP]	12.78	5.6	0.28	6.6	0.47	5.1	Terminal GG helix junction
Tl^+ 106	12	14.70	6.6	0.32	5.5	0.39	7.3	TpG step minor grove
Rb^+ 101[SP]	15[SP]	8.73	18.0	0.38	26.2	0.39	14.5	Terminal GG helix junction
Rb^+ 102	18	11.19	8.6	0.14	12.7	0.40	8.1	Terminal GG major groove

[a]*Isotropic Thermal factor (see chapter 1)*
[b]*Peak height of anomalous Fourier electron density*
[c]*Distance between $|F_+-F_-|$ peak and the position of the refined Tl_+/Rb_+ ion*
[d]*Peak height of Isomorphous Fourier electron density*
[e]*Distance between $F_{Tl}/F_{Rb}-F_{Na}|$ peak and the position of the refined Tl_+/Rb_+ ion*
[f]*Residual difference density after fully occupied water was modeled at the site*
[SP]*Atoms are on crystallographic special position, which is on 2-fold axis. Occupancy is doubled when considering a biological unit (asymmetric unit is a single strand)*

$[d(CCAGGCCTGG)]_2/Na^+$ structure (PDB ID 3GGB, see chapter 2). Hence it is not biased by positions of heavy atoms.

The positions of heavy atoms determined in the current structures are refined against corresponding $2F_o-F_c$ and F_o-F_c peaks, thus not affected by the either anomalous Fouriers or isomorphous Foriers. However, in any case, the appearance of both anomalous and isomorphous fouriers are consistent with the position of refined heavy atom sites. Additionally, the assignment of monovalent cations were held until the refinement with water molecules was at near-convergence, including all atomic positions,

thermal factors and phases. Hence the model bias toward heavy atoms during the refinement process is effectively avoided.

Refinement

A total of 13522 reflections in the resolution range 15.3-0.98 were used in the refinement for $[d(CCAGGCCTGG)]_2/Tl^+$ and 17543 reflections in the resolution range 30.60-0.87 were used in $[d(CCAGGCCTGG)]_2/Rb^+$. In both cases, there is no exclusion of the reflections by sigma-cut, but by the limitation of the maximum resolution used due to the low completeness of the resolution bin at higher resolution. The final *R factor* is determined to be 12.02% (Tl^+) and 11.97% (Rb^+) for all data. The asymmetric unit contains single strand of d(CCAGGCCTGG). As customary, bases are number 1-10 in the first strand, 11-20 in the other. Two strands are related by the crystallographic 2-fold axis of C2 space group to complete a biological unit. All atoms for DNA have been modeled in and coordinates are well defined, as indicated by the clean and defined electron density maps (Figure 3.1-3.2) and low thermal factors determined (<10 for DNA coordinates, table 3.1). As with previous $d(CCAGGCCTGG)]_2/Na^+$ structure (see chapter 2), both final models include frequently appeared disordered regions mainly in its sugar/phosphate backbone region, but also extends to the base pairs at the center of duplex. Close inspection of both $2F_o$-F_c and F_o-F_c electron densities also revealed five partially occupied hydrated (hexaaqua) Mg^{2+} ions in both decamer structures (per asymmetric unit). The locations of these disordered and/or partially occupied Mg^{2+} complexes are nearly identical to those in previous $d(CCAGGCCTGG)]_2/Na^+$ structure, and number 21-25. Total of 76 (DNA/Tl^+) and 91 (DNA/Rb^+) water molecules (including partially occupied) per asymmetric unit were added to the final refined model.

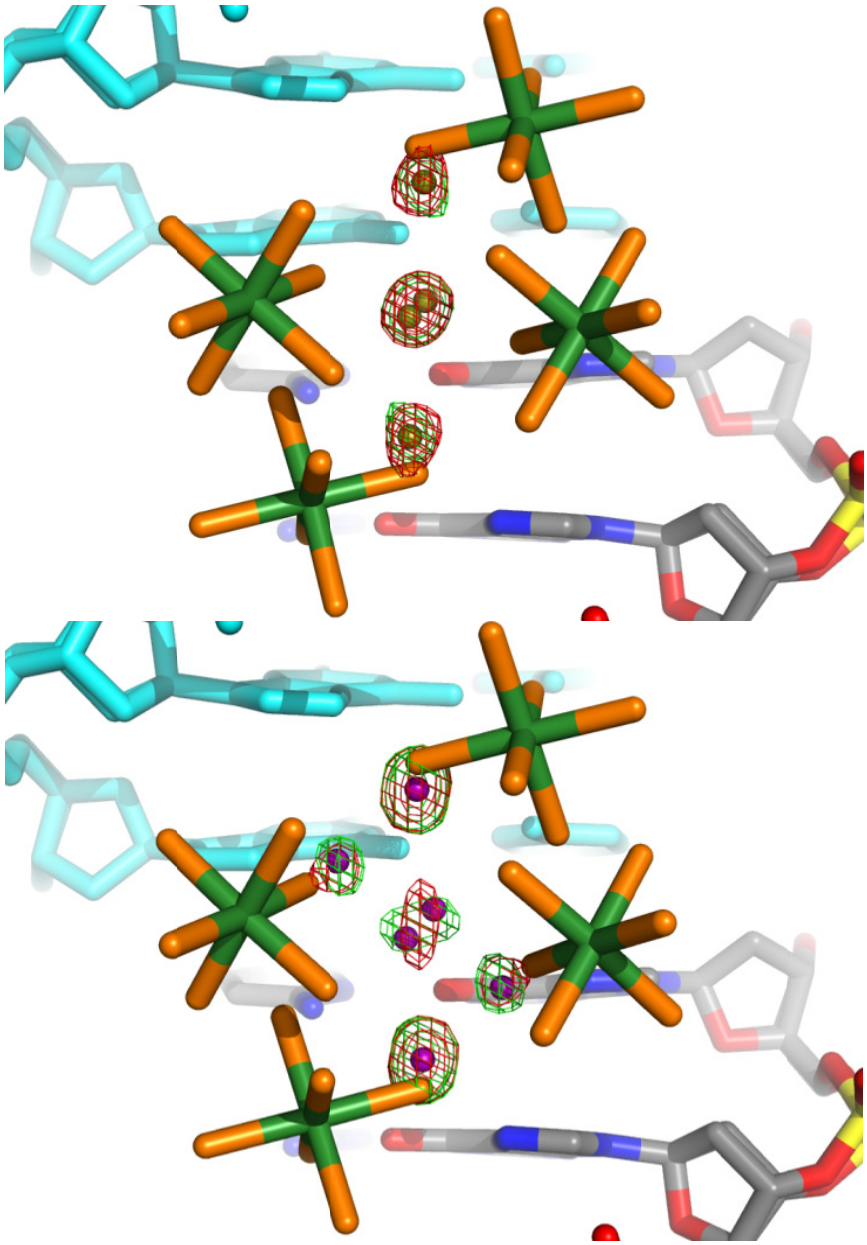

Figure 3.4: Comparison of anomalous and isomorphous peaks at the terminal GG/CC step - helical junction in the crystal of $CCAGGCCTGG_2$. In crystal space group C2, helices are stacked on top of each other through crystal lattice, forming nearly continuous strands of double helix. Rb^+ identified only at this location in this work. Anomalous and isomorphous difference densities are shown in green and red, respectively. Both are contoured at 3.5σ. (Top) Rb^+ 101-102 with surrounding Mg^{2+} complexes 22 (alternate states a & b). (Bottom) Tl^+ 105, 104 and 101 with Mg^{2+} 22a and b. Crystallographic 2-fold symmetry axis runs at the center of each figure, where Rb^+ 101 and Tl^+ 105 are located (crystallographic special position). Slight offset creates a set of closely located partially occupied ions (they do not present at indicated positions at the same time. Interactions of these cations to DNA are described in the discussion section of this chapter.

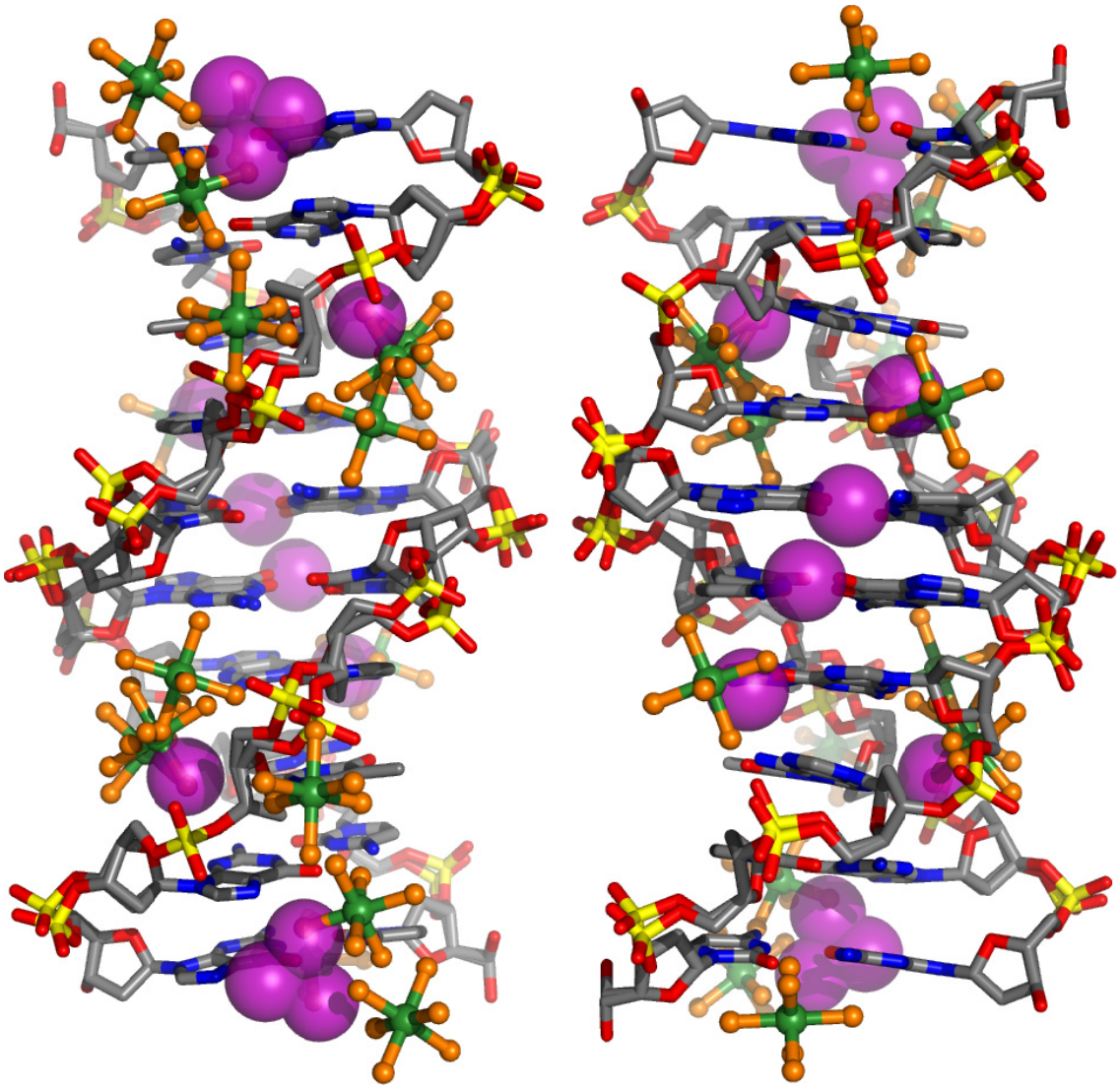

Figure 3.5: Overall representation of localized cations and DNA determined from the crystal of [d(CCAGGCCTGG)]$_2$/Tl$^+$. Viewing into major groove (left) and minor groove (right). Some hexaaquaMg^{2+} ions are highly disordered also observed in the crystal structure of Na$^+$ form (see chapter 2) and Rb$^+$ (this work) of the same oligomer. DNA strands are shown in stick models, colored with CPK coloring scheme. Thallium (I) ions are shown in purple spheres. Each Mg ion is shown with ball and stick, colored green at its metal center and orange for the coordinated water molecules. Water molecules are omitted for clarity.

e.

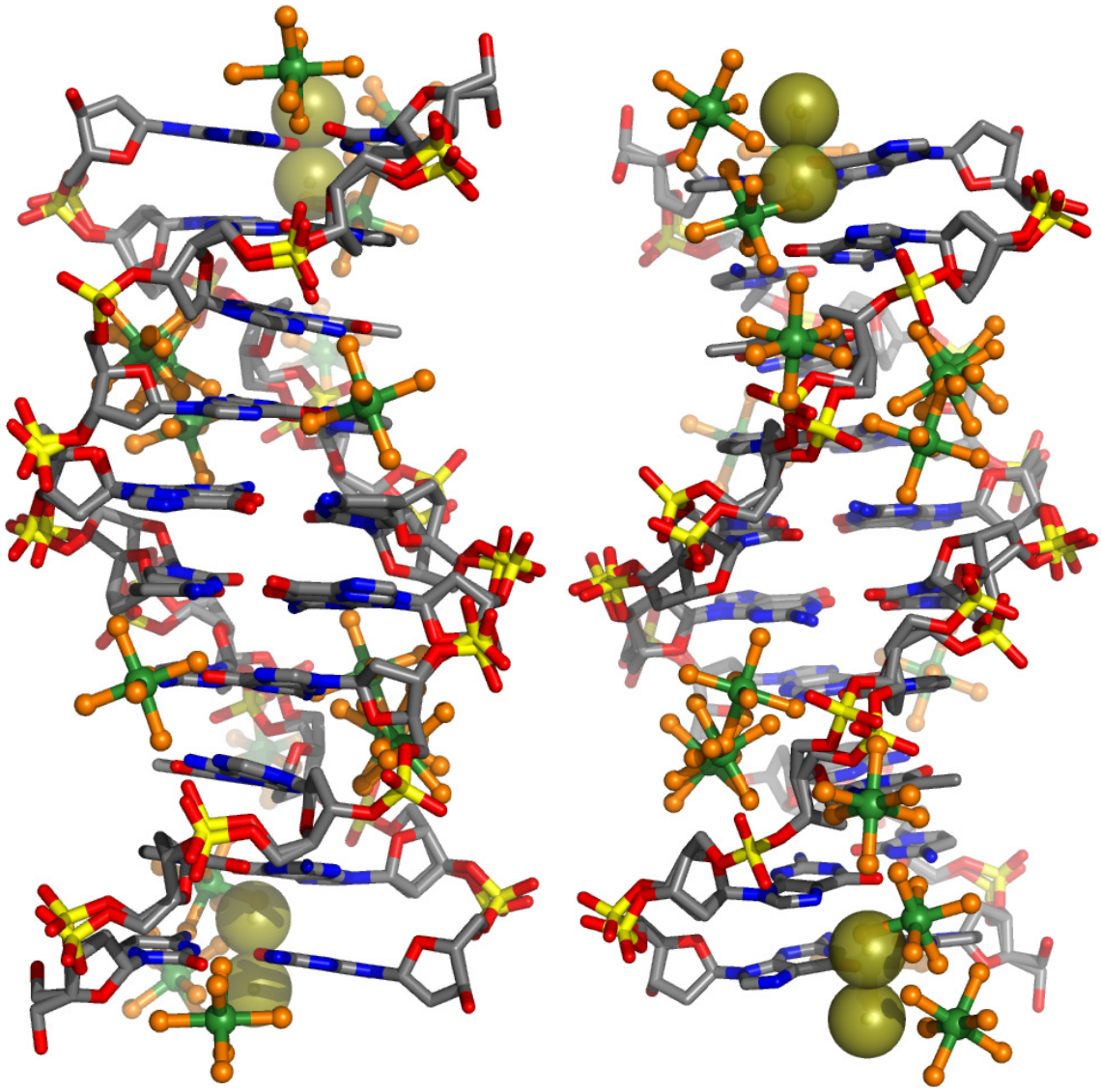

Figure 3.6: Overall representation of localized cations and DNA determined from the crystal of [d(CCAGGCCTGG)]$_2$/Rb$^+$ Viewing into major groove. (Left) and minor groove (Right). Some hexaaqua Mg^{2+} ions are highly disordered also observed in the crystal structure of Na$^+$ form of the same oligomer (see chapter 2). DNA strands are shown in stick models, colored with CPK coloring scheme. Rubidium (I) ions are shown in olive spheres. Each Mg ion is shown with ball and stick, colored green at its metal center and orange for the coordinated water molecules. Water molecules are omitted for clarity

Twelve partially occupied Tl$^+$ ions (six are symmetry related) were added to the final refined model for [d(CCAGGCCTGG)]$_2$/Tl$^+$. Four partially occupied Rb$^+$ ions (two

are symmetry related) were added to $[d(CCAGGCCTGG)]_2/Rb^+$. Positions of Rb^+ ions in the refined model are consistent with two Tl^+ positions identified in this work (figure 3.4). In both Tl^+ and Rb^+ cases, strong residual F_o-F_c Fourier electron density superimposed on the $2F_o$-F_c Fourier density of a refined water molecule (figure 3.4). Thus it was used to initially identify heavy atom sites as it indicates poor fit of the model and the necessity for increased scattering possibly by heavy atoms. The residual F_o-F_c peak heights measured at these sites are considerably higher than the criterion previously used by our group. Howerton et al. used a 3.5σ criterion for initially identifying species ranging from 8 electrons (10% occupied Tl+ ions) to 28 electrons (35% occupied Tl+ ions) in F_o-F_c Fourier electron density map. In $[d(CCAGGCCTGG)]_2/Tl^+$, the residual Fo-Fc Fourier electron density peaks observed here range from 4.5σ (6% occupied Tl^+ ion) to 29.3σ (24% occupied Tl^+ ion). In $[d(CCAGGCCTGG)]_2/Rb^+$, residual peaks range from 8.1σ (18% occupied Rb^+ ion) to 14.5σ (30% occupied Rb^+ ion).Hence our criteria used for initial screening of heavy atom sites here is more stringent, as no site with residual F_o-F_c peak below 4.5σ was included.

Occupancies of the Tl^+ and Rb^+ atoms were initially estimated with the refinement program *SHELXL* (Sheldrick 2008) as with other regions where multiple states are present. During the automated occupancy refinement, additional restrains were applied to occupancies at where multiplicity of the site (occupied by two or more atoms) was observed to keep the sum of the occupancies below unity. Automated occupancy refinement was no mean to finalize the estimated occupancies for these heavy atom sites, as occupancy for each site was carefully adjusted by inspecting appearance of corresponding $2F_o$-F_c and F_o-F_c peaks. The assignment of heavy atoms was carried from

the site where highest peaks of F_o-F_c, F_+-F_- and F_{Tl+}-F_{Na+} or F_{Rb+}-F_{Na+} were observed. The trend is generally consistent with that of occupancy displayed by these heavy atoms and most of them are located within the major groove. The only minor-groove bound monovalent cation was identified with Tl^+ data, and it was assigned at last. When the refinement was near convergence, the F_o-F_c peak at height of 7.3σ surrounded the site, which is shared with first shell water molecules of two Mg^{2+} complexes (a total of 45% occupancy). The site was replaced with 12%/45% Tl^+/water hybrid (Tl^+ 106), effectively diminishing the F_o-F_c peak as the number of electrons was increased from 4.5 to 14.

Isomorphous difference electron density Fourier

High isomorphism of the DNA coordinates between $[d(CCAGGCCTGG)]_2/Tl^+$ and $[d(CCAGGCCTGG)]_2/Na^+$ or $[d(CCAGGCCTGG)]_2/Rb^+$ and $[d(CCAGGCCTGG)]_2/Na^+$ crystals are clearly indicated by low R_{iso} upon merging data (0.203/0.137), negligible difference between unit cell parameters (max Δ=0.4%/0.1%) and RMS deviations (0.038Å/0.047Å). It is feasible to identify locations of heavy atom sites based on F_{Tl+}-F_{Na+} or F_{Rb+}-F_{Na+} difference Fourier, as isomorphous peaks may arise from heavy atom scatterings that do not present in DNA/Na^+ crystal. However, in this work, no heavy atom site was yet identified solely by the presence of isomorphous peak as they may also incorporate other origins.

The possible contribution to isomorphous peaks includes difference in locations and/or occupancies of localized water and cations. In this work, the appearance of negative and positive F_{Tl+}-F_{Na+} or F_{Rb+}-F_{Na+} difference peaks are apparent around alternating hexaaqua Mg^+ complexes. These isomorphous peaks originate from slight difference in observed occupancies of these disordered Mg^{2+} complexes, hence resulting

in different scatterings between the given set of data. Partially dehydrated monovalent cations and hydrated divalent cations are known to compete for a common binding region in the G-tract major groove (Shui, McFail-Isom et al. 1998; Tereshko, Minasov et al. 1999; Tereshko, Minasov et al. 1999; Sines, McFail-Isom et al. 2000; Woods, McFail-Isom et al. 2000; Howerton, Sines et al. 2001). Several Tl^+ and Rb^+ ions in this work are also observed to display similar results. Hence the use of F_{Tl+}-F_{Na+} or F_{Rb+}-F_{Na+} difference peaks alone potentially results in erroneous identification of heavy atom sites in this case. The use of isomorphous peak was always performed in conjunction with anomalous difference Fourier.

Anomalous difference electron density Fourier

Anomalous $|F_+$-$F_-|$ difference Fourier can be used more reliably than isomorphous difference Fourier to locate heavy atom sites in general. Anomalous $|F_+$-$F_-|$ difference Fourier arise from differences between Bijvoet pairs due to anomalous scattering of heavy atoms. However, not all $|F_+$-$F_-|$ peaks do not warrant the position of the heavy atom sites, as it is apparent from several observed $|F_+$-$F_-|$ peaks that are inconsistent with appearance of $2F_o$-F_c and F_o-F_c peaks. Hence another criterion was imposed when employing anomalous difference Fouriers (criterion c). In $[d(CCAGGCCTGG)]_2/Tl^+$, two symmetry related $|F_+$-$F_-|$ peaks with the height at 4.5σ appear in the minor groove, near central G-tract. The peak is 1.6Å from HOH224, 2.7Å from O2 of C(6) and 2.3Å from N2 of G(15), both in DNA state A. Due to its inconsistency with sum/difference peak and too-short contact to the polar group of guanine, this $|F_+$-$F_-|$ peak is hence considered as a noise. Since it is the highest ranked peak among observed noise, the peak height of 4.5σ is thus used as a cut-off when considering a signal-to-noise ratio of anomalous scattering

in $[d(CCAGGCCTGG)]_2/Tl^+$ data (one of the criteria as listed in materials and methods) . Similarly, minimum peak height of 4.2σ is defined for $[d(CCAGGCCTGG)]_2/Rb^+$ data.

It is possible that the heavy atom site indicated by $|F_+-F_-|$ peak may not be consistent with the presence of $F_{Tl+}-F_{Na+}$ or $F_{Rb+}-F_{Na+}$ peak. The possible scenario includes when there is no difference in estimated number of electrons occupying the same site between native and derivative data set. For example, consider the case when 12% occupied Tl^+ is present in DNA/Tl^+ crystal, and water fully occupies the same site in DNA/Na^+ crystal. In this case, $|F_+-F_-|$ peak would indicate the presence of Tl+, however $F_{Tl+}-F_{Na+}$ peak would not appear at the site due to the same number of electrons (10 *Coulombs*) contributed by each species. In the current work, with minimum height requirement for anomalous peaks, no heavy atom site with anomalous peak lacks corresponding isomorphous peak (except for the phosphorous of DNA due to weak anomalous scattering of P). The lowest occupied Tl^+ 104 (6% occupied) site still shows difference in number of contributing electrons as the exact site is not occupied by water in $[d(CCAGGCCTGG)]_2/Na^+$ crystal. Similarly, all other low occupied heavy atom sites show either difference in total occupancies or position (0.4-0.7Å) when compared to corresponding water molecules in $[d(CCAGGCCTGG)]_2/Na^+$ crystal. Hence all heavy atom sites identified in this work incorporate both $|F_+-F_-|$ peak and $F_{Tl+}-F_{Na+}$ or $F_{Rb+}-F_{Na+}$ peak at the respective position.

Verification of heavy atom sites

Verification of the Tl^+ and Rb^+ positions in the refined models were performed with both $|F_+-F_-|$and $F_{Tl+}-F_{Na+}$ or $F_{Rb+}-F_{Na+}$ Fouriers. No refinement incorporates information or assumption about the positions or scattering of heavy atom from either

anomalous or isomorphous difference Fouriers. Both refined structures of $[d(CCAGGCCTGG)]_2/Tl^+$ and $[d(CCAGGCCTGG)]_2/Rb^+$ and are consistent with the $|F_+$-$F_-|$and F_{Tl+}-F_{Na+} or F_{Rb+}-F_{Na+} Fouriers. All refined Tl^+ and Rb^+ sites are confirmed by both $|F_+$-$F_-|$and F_{Tl+}-F_{Na+} or F_{Rb+}-F_{Na+} Fouriers.

Discussion

Localized Tl^+ and Rb^+ ions

Twelve Tl^+ ions (six are symmetry related) were added to the final refined model for $[d(CCAGGCCTGG)]_2/Tl^+$. Four Rb^+ ions (two are symmetry related) were added to $[d(CCAGGCCTGG)]_2/Rb^+$. Positions of Rb^+ ions in the refined model are consistent with two Tl^+ positions identified in this work. Both Tl^+ and Rb^+ positions are predominantly within the grooves, and in general are not in direct proximity to phosphate groups. The majority of the observed Tl^+ and Rb^+ ions interact with a single duplex; therefore they are not engaged in lattice interactions or crystal packing. None of the observed Tl^+ or Rb^+ is fully occupied. The most highly occupied sites are located within the major-groove of CGCGAATTCGCG, with estimated occupancies ranging from 6-24% (Tl^+) and 18-30% (Rb^+). Two 2-fold symmetry related Tl^+ ions are localized within the minor-groove with estimated occupancy of 6% (figure 3.5-3.6).

Providing high frequency of disordered and/or partially occupied hexaaqua Mg^{2+} complexes surrounding the duplex of CCAGGCCTGG, Tl^+ and Rb^+ sites identified in this work often compete for the common binding site within both the major and the minor grooves. Both Tl^+ ions in the minor groove, four Tl^+ ions and two Rb^+ ions in the major groove occupy ~2.5Å from metal center of adjacent Mg^{2+} complexes, nearly overlap with one of the first hydration shell water molecules of a given Mg^{2+} complex.

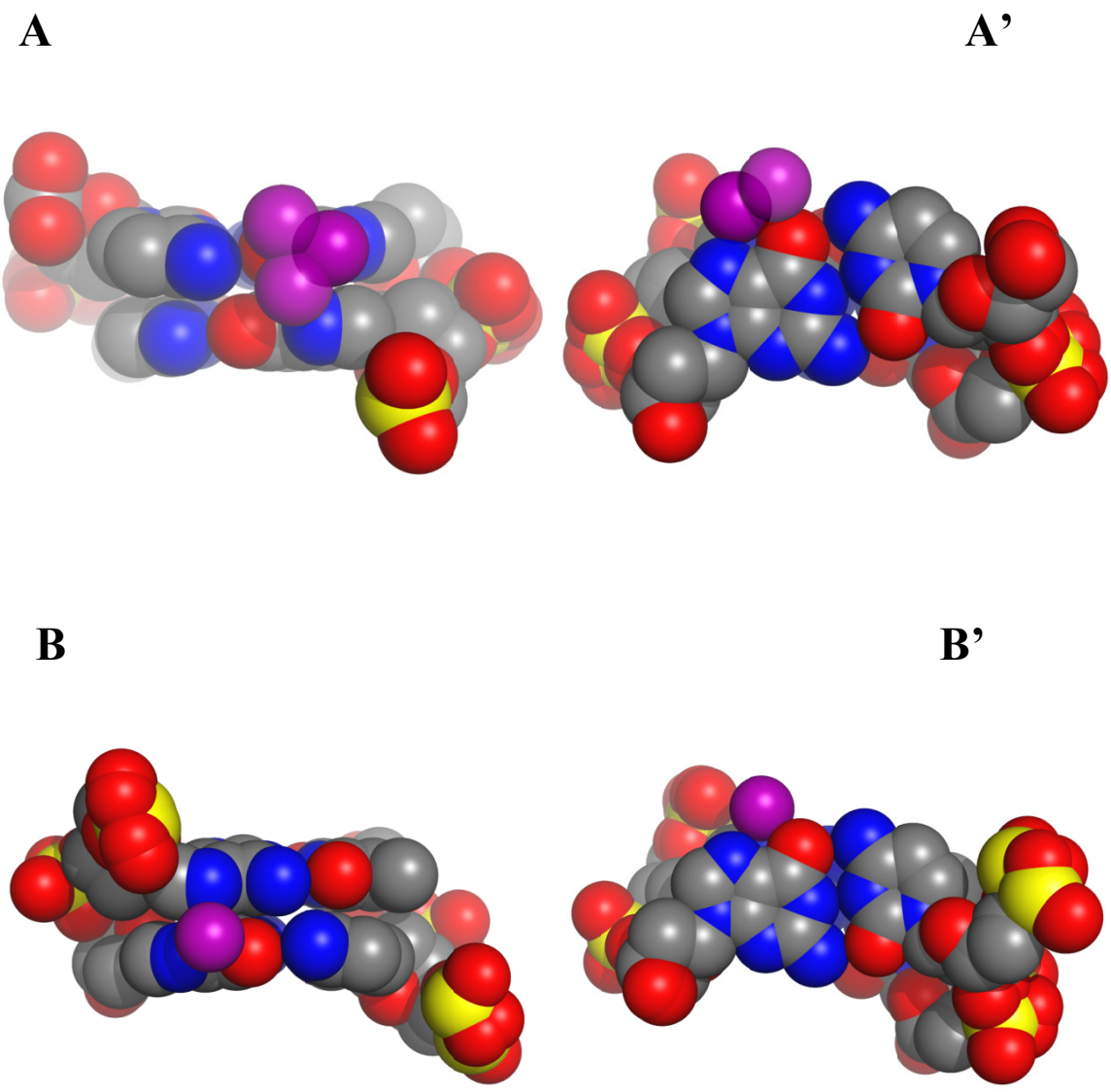

Figure 3.7a: Groove interactions of Tl^+ with adjacent residues. For description of the figures, see the legend for figure 3.7b.

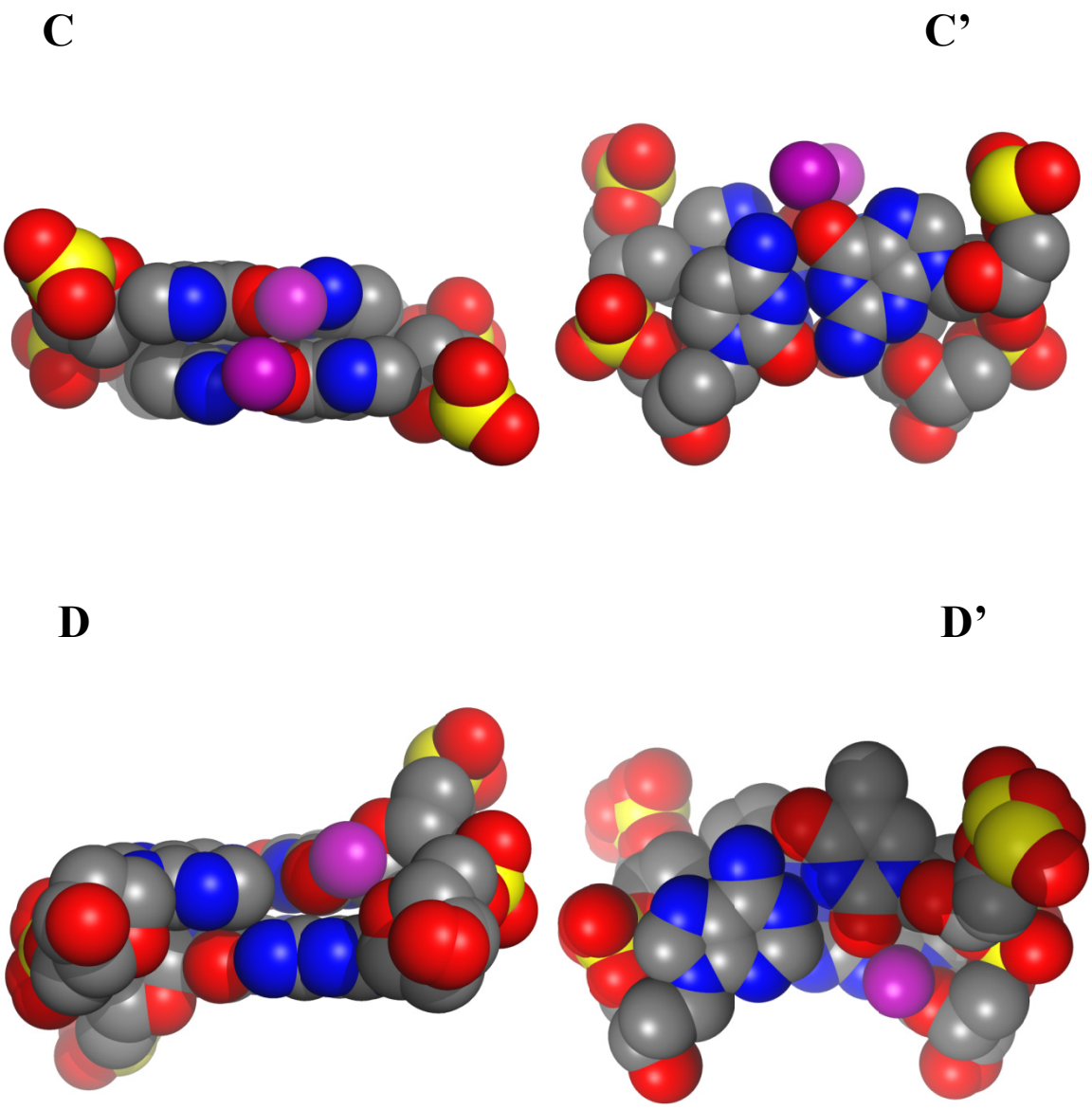

Figure 3.7b: Groove interactions of Tl^+ with adjacent residues. Parallel (left) and lateral (right) with respect to the plane of a given base pairs are shown in each set of figure. The right figure (eg. A') always shows the major groove at top, and the minor groove at bottom. Tl^+ ions are shown in purple. (A/A') Terminal GG/CC step with three partially occupied thallium ions, from the top, 105, 104 and 101 are shown. (B/B') Tl^+ 102 at the ApG step (C/C') Two Tl^+ ions, 103 at the central G-tract of the CCAGGCCTGG. In this structure, two strands of DNA are related by a 2-fold symmetry axis that runs at the center of duplex, perpendicular to the helical axis (asymmetric unit is a single strand). (D/D') Tl^+ 106 bound at the TpG step in the minor groove. Tl^+ 106 interacts with both thymine O2 and O4' of deoxyribose of guanine. Specific interactions including bonding distances are explained in the discussion section of this chapter.

In any case, combined estimated occupancy between Tl^+/Rb^+ and its competing Mg^{2+} complex does not exceed 100%, ranging from 64% to 100%. Similar monovalent /divalent cation competition for a common site was previously reported for a catalytic RNA (Basu, Rambo et al. 1998) and the G-tract major groove of CGCGAATTCGCG (Howerton, Sines et al. 2001).

Involvement of hydrated divalent cations in axial bent across the non-bonded junction between stacked helices of *B*-DNA has been previously suggested, including CCAAGATTGG/Mg^{2+}, CCAAGCTTGG/Ca^{2+}, CCAGCGCTGG/Mg^{2+} and CCAGCGCTGG/Ca^{2+} (Chiu and Dickerson 2000). The observed junction-bent between stacked helices appear to be the result of Mg^{2+}/Ca^{2+}- induced consecutive positive roll angles at terminal base pairs toward the major groove, which is also observed in the current structure (see chapter 3). Mg^{2+} complexes in both the major and the minor grooves at the CA/TG step also appear to correlate with its unusual base-pair stacking, which was originally reported by Heinemann and Alings (Heinemann and Alings 1989) (see chapter 3). The observation here that the same sites are shared with partially occupied Mg^{2+} complex and Tl^+/Rb^+ ions may thus suggest not fully, but partial contribution to the deformation of *B*-DNA.

Cations in the major groove

Nearly all Tl^+ and Rb^+ sites identified in this work (except Tl^+ 106) are located within the major groove. In essence these Tl^+ and Rb^+ positions are in accord with previous studies (Howerton, Sines et al. 2001; Howerton, Nagpal et al. 2003). Both refined Tl^+ and Rb^+ positions in the major groove are in proximity to adjacent guanine bases (figure 3.7, 3.9). A total of five out of six Tl^+ ions and two Rb^+ ions (both per

asymmetric unit) are within 3.5Å of O6 of guanine. A total of four out of six Tl^+ ion and one Rb^+ ion (both per asymmetric unit) are within 3.5Å of N7 of guanine. Providing G/C rich sequence of DNA used in this study, O6 of all guanine has Tl^+ ion within 3.5Å in $[d(CCAGGCCTGG)]_2/Tl^+$, except of G(9) and G(19) where the adjacent Tl^+ 101 is 3.8Å away.

In the G-tract major groove, geometry displayed by basic functional groups (O6 and N7 of guanine) affects the position of monovalent cations (Howerton, Sines et al. 2001). In their work, localized Tl^+ in the G-tract major groove of a duplex CGCGAATTCGCG are observed predominantly to have bidentate chelation with both N7 and O6 of a single guanine. Providing a G/C rich sequence and a distinctively different patter of the G-tract in the current study (CCAGGCCTGG), local variation of the major groove Tl^+ positions with respect to the adjacent guanine is apparent. Only four out of ten Tl^+ localized in the major groove (102 and 104) are chelated in the same manner. Instead, intra-strand chelation by the N7 of preceding purine is observed at all purine-purine steps (ApG and GpG). Hence, the observed additional intra-strand chelation is consistent. There is no ApG or GpG step in the duplex of CGCGAATTCGCG, so is no such intra-strand chelation by the N7 of preceding purine in previous work. This intra-strand chelation by N7 appears to affect parallel and lateral position of the Tl^+ and Rb^+, with respect to the plane of adjacent guanine. This may thus explains the local contact variations observed, with respect to both N7 and O6 of the adjacent guanine residue. Providing frequent purine-guanine steps in the duplex of CCAGGCCTGG and consistent results, monovalent cations appear to localize adjacent to *B*-DNA bases not only in site-specific, but also in sequence-dependent manner.

Cations in the G-tract major groove (Tl+ 102 and 103)

A total of four Tl^+ sites (Tl^+ ions 102-103 and their symmetry related pairs) are located in the central G-tract major groove, GGCC (figure 3.8). The estimated occupancies are 11% for Tl^+ 102 and 16% for Tl^+ 103. Additional $|F_+-F_-|$ peak (4.1σ) and $F_{Tl+}-F_{Na+}$ peak (5.0σ) appear in proximity to Tl^+ 103. In the final refined model, a partially occupied water molecule is assigned to the site (HOH 238c, 35%) as the observed height of $|F_+-F_-|$ peak does not satisfy one of the selection criteria. The site may be suggestive as an additional Tl^+ site with the estimated occupancy of 4%, possibly a disordered part of Tl^+ 103 (Table ###). Tl^+ 102 site is slightly shifted from the plane of adjacent G(4) toward preceding A(3). The same ion shares the site with one of first shell water molecules of Mg^{2+} 21a, which appears to show a correlation with the unusual base stacking at the CA/TG step with large helical twist (see chapter 3). Whereas Tl^+ 103 positions nearly within the plane of adjacent G(5). Tl^+ 102 site is with3.0Å from both N7 and O6 of G(4), whereas Tl^+ 103 is in comparison shifted toward the center of the groove. Tl^+ 103 is chelated by O6 but not by N7 of G(5). Instead, Tl^+ 103 is chelated by inter-strand O6 of G(15b).

Both Tl^+ sites are adjacent to the guanine at "purine-guanine (Tl^+) step", where the preceding purine is A(3) or G(4). Both Tl+ 102 and 103 are additionally chelated by N7 of the preceding purine within the distance of 3.4Å. This additional inter-strand chelation by N7 of preceding purine appears to correlate with the parallel position of Tl^+ ions with respect to the plane of adjacent guanine. The positions of respective N7 and G(4) and G(5) is distinctively different. At the ApG step, the ideal base-base stacking is nearly conserved. The smallest twist angle of ~20° seen in the duplex situates N7 of A(3)

atop of a site between N7 and O6 of G(4). Whereas increased twist by ~15° at the GpG step reduces the base-base stacking only to the 6-membered ring of both purines. N7 of G(4) lies further closer to atop of O6 of G(5). The positioning of preceding purine N7 thus appears to correlate with respective Tl^+ site, as Tl^+ 102 is positioned between N7 and O6 of G(4) within 3.0Å, whereas Tl^+ 103 is positioned toward O6, but further away from N7 of G(5).

Cations at the terminal GG/CC step (Tl+ 101 104 105 & Rb+ 101 102)

A cluster of three Tl^+ sites (101, 104 and 105) are located at the both terminal GG/CC step in $[d(CCAGGCCTGG)]_2/Tl^+$. All Rb^+ sites identified in this work are located at the same terminal GG/CC step in $d(CCAGGCCTGG)]_2/Rb^+$(figure 3.9). Positions of Rb^+ ions are consistent with two Tl^+ positions, as Rb^+ 101/Tl^+ 105 and Rb^+ 102/Tl^+ 101 nearly superimpose each other. The estimated occupancies are 6-24% for Tl^+ ions and 18-30% for Rb^+ ions at this site.

Tl^+ 104 lies on the plane of terminal G/C base pair. Here, the same purine-guanine (Tl^+) step rule mentioned in the previous section applies. The twist angle of the GpG step is reduced at the terminal of the duplex, with similar angle found in the ApG step. Hence, the pattern of Tl^+ chelation is the same between both steps. In addition to bidentate chelation by both N7 and O6 of G(10), the geometry of basic functional groups at this GpG step allows an additional contact by intra-strand N7 from preceding G(9). With its low occupancy (6%) and close location to both Tl^+ 101 and 105 (< 2.2Å), Tl^+ 104 is assumed to be a disordered part of either of the neighboring Tl^+ ions. On the other hand, Tl^+ 101 site is situated at the terminal GpG step, between planes of adjacent base pairs. Intra-strand chelation from N7 of G(9) and O6 of G(10) is observed.

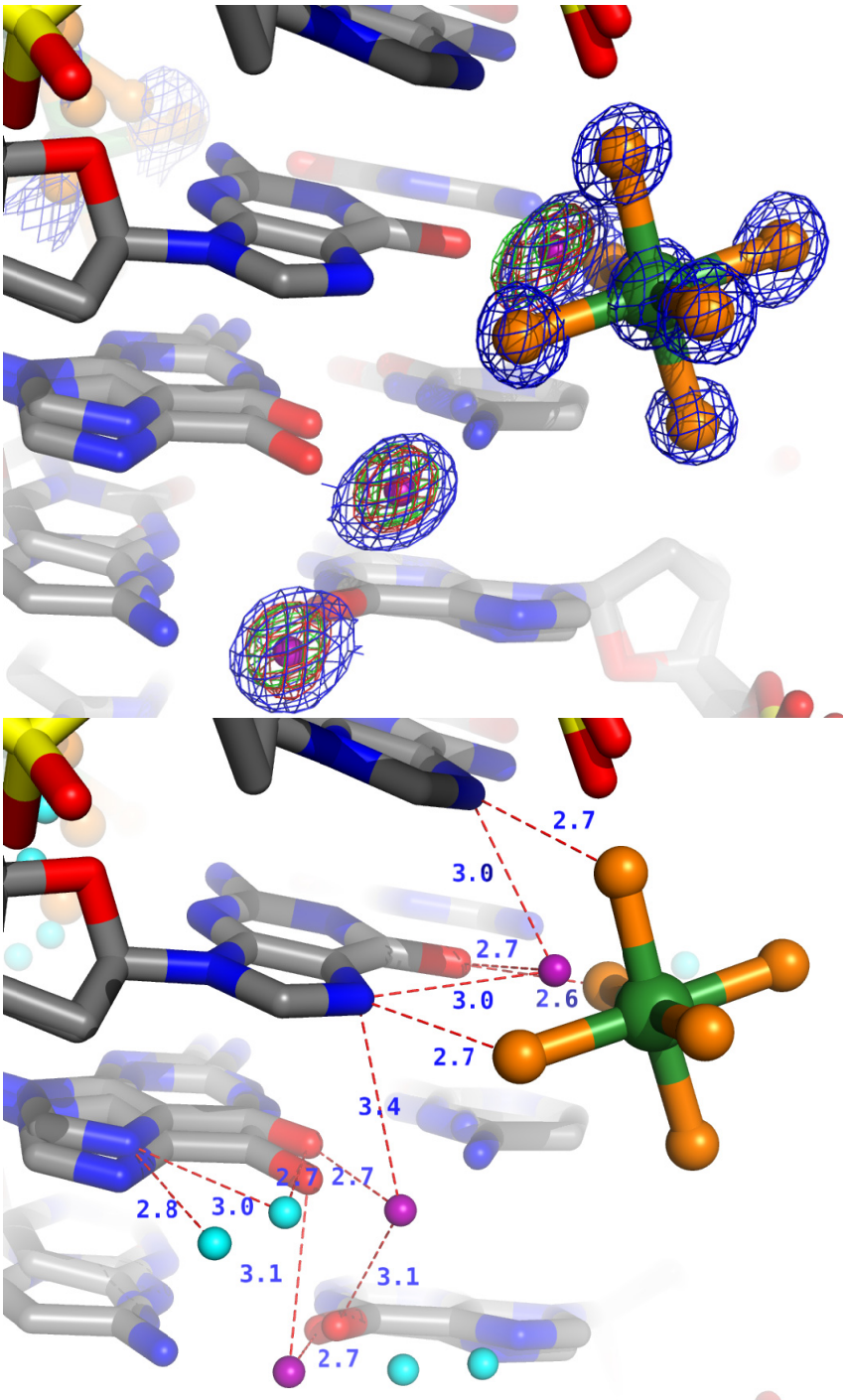

Figure 3.8: G-tract major groove interactions of Tl^{+} ions 102 (top of the figure) and 103 (bottom 2), Mg^{2+} 21 with adjacent residues in $[d(CCAGGCCTGG)]_2$. (Top) Shown with sum difference Fouriers (Fo-Fc) in blue, contoured at 1σ around cations. Anomalous difference densities (green) and isomorphous difference densities (red) are also shown, both contoured at 3.5σ. Notice the close location of Tl^{+} 102 (at the top of the figure) to Mg^{2+} 21,which indicates the possible competition for a common binding sites between these cations. Estimated occupancy for Mg^{2+} 21 is 85%, and 11% for Tl^{+} 102. Tl^{+} 103 (two at the bottom) are 16% occupied and they are 3Å apart. Bottom figure shows contact distances in the unit of angstrom, as well as HOH 238a and 238b that are possibly partially occupied Tl^{+} ions due to the weak anomalous and isomorphous difference peaks observed at the site. See text for explanation.

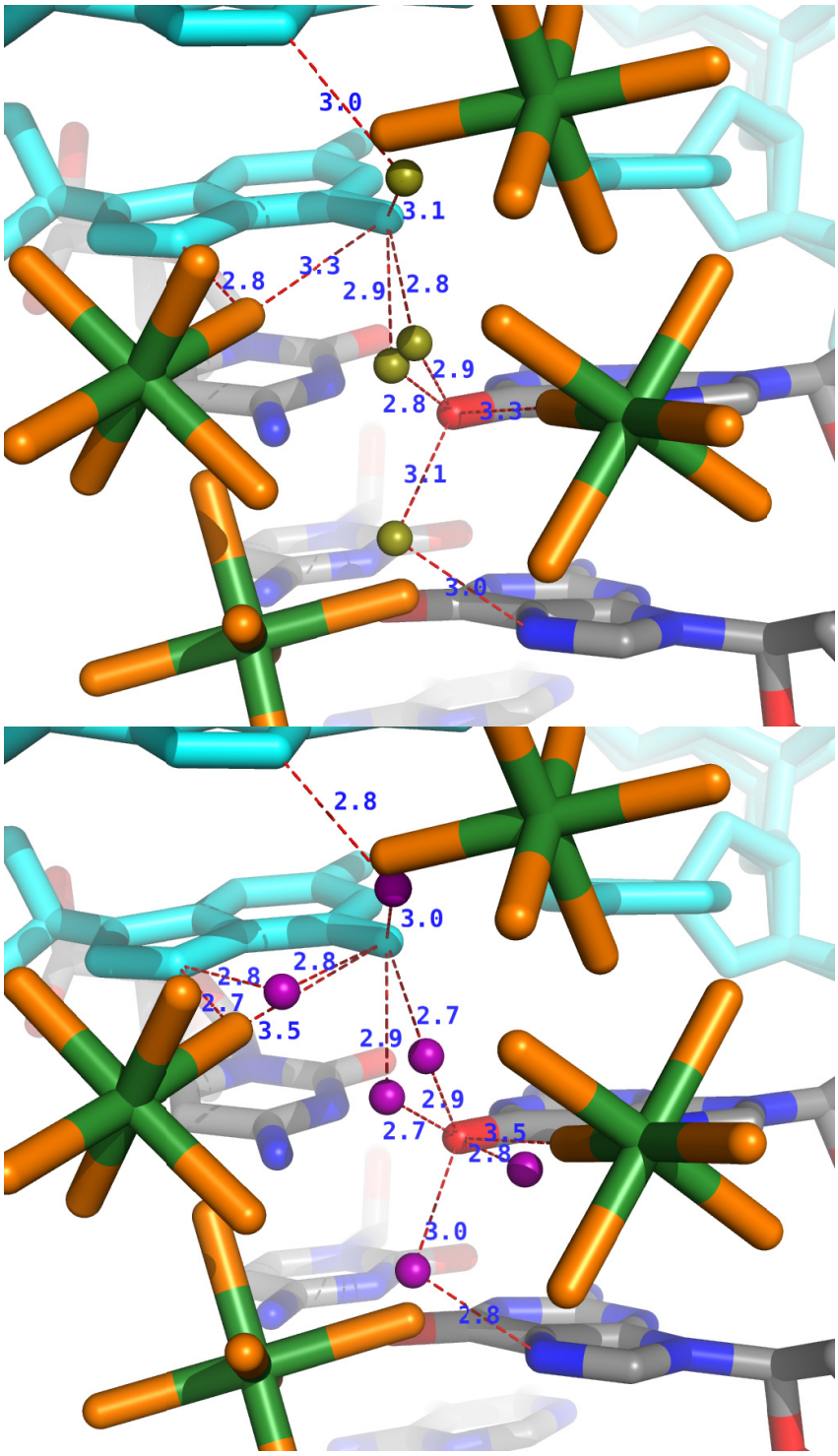

Figure 3.9: Interactions of Rb^+ and Tl^+ ions at the terminal GG/CC step-helical junction in the crystal of $CCAGGCCTGG_2$. Their binding sites are nearly consistent, including interactions with surrounding molecules. (Left) Rb^+ 101-102 with surrounding Mg^{2+} complexes 22 (alternate states a & b). Rb^+ ions are shown in olive-green sphere, (Right) Tl^+ 105, 104 and 101 with Mg^{2+} 22a and b. Crystallographic 2-fold symmetry axis runs at the center of each figure, where Rb+ 101 and Tl+ 105 are located (crystallographic special position). Slight offset creates a set of closely located partially occupied ions (they do not present at indicated positions at the same time).

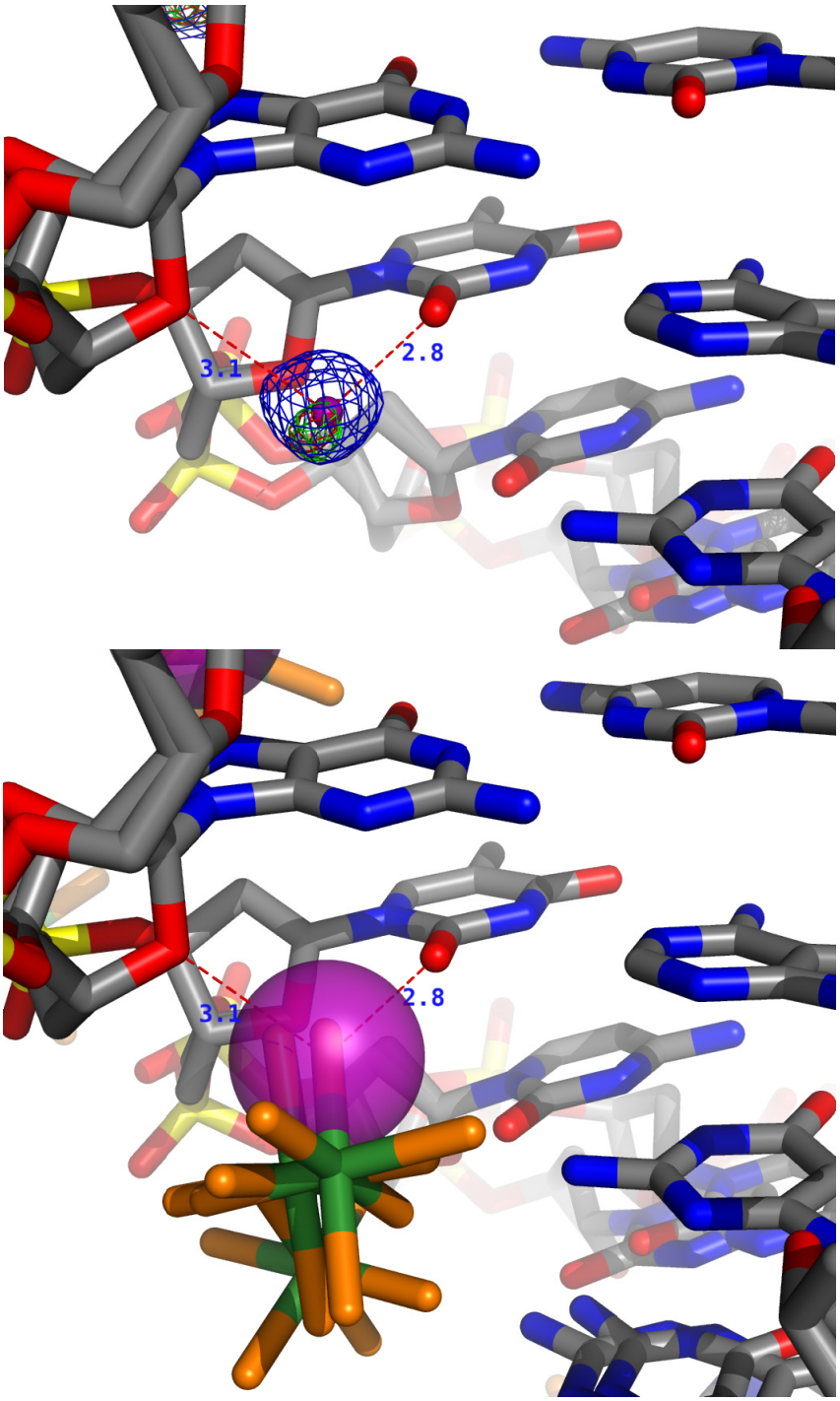

Figure 3.10: Minor-groove bound Tl^+ 106 at the TpG step. Tl^+ 106 interact with thymine O2 and O4' of deoxyribose of G(9). This is the only observed monovalent cation biding in the minor groove of $CCAGGCCTGG_2$. (Left) Shown with sum difference Fouriers (F_o-F_c) in blue, contoured at 1σ around cations. Anomalous difference densities (green) and isomorphous difference densities (red) are also shown, both contoured at 3.5σ.See table 2.1 for peak heights. (Right) Tl^+ 106 (6% occupied) is shown in transparent sphere, colored purple, and Mg^{2+} 24 with three alternating states, a (35%), b (15%) and c (30%). The position of Tl^+ 106 overlaps with first hydration shell of Mg^{2+} in state b and c and 2.4Å from that of Mg^{2+} 22a. Hence the site is nearly occupied fully (totaling 96%).

The same ion shares the site with one of first shell water molecules of Mg^{2+} 22a, which appears to show a correlation with positive roll angles at the terminal CC/GG step (see chapter 2).

Tl^+ 105 (Rb^+ 101) site is located on a crystallographic 2-fold special position at the junction created by two helices that are stacked top to bottom. Therefore, each of Tl^+ 105 and Rb^+ 101 is located between the planes of terminal G/C pair and first C/G pair among stacked helices. Both ions are chelated by inter-strand and inter-helix O6 atoms of guanine residues. In crystal space group *C2*, dyad symmetric helices stack along the *c*-axis of the unit cell to form layers of continuous helices that propagate through crystal lattice. The step from the last base-pair of one helix to the first base-pair of the next helix along a given column is therefore similar to the GpC within helices. This G-junction-G step nearly preserves local geometry found at the GpC step in the central G-tract of CCAGGCCTGG, such as base-pair stacking. Tl^+ 105 (Rb^+ 101) site and its chelation are thus similar to between-plane Tl^+ ion located at the GpC step of the duplex CGCGAATTCGCG (Tl^+ 2107 and 2113), which is also chelated by two inter-strand O6 atoms (Howerton, Sines et al. 2001).

Cations in the minor groove

The minor groove of the duplex of CCAGGCCTGG contains two Tl^+ sites that are related by crystallographic 2-fold symmetry (figure 3.10). Each site is located at the TpG step, and is partially occupied by Tl^+ 106 (6%). This partially occupied Tl^+ site is also shared with first shell water molecules of disordered Mg^{2+} 24b (30%) and 24c (15%). The Tl^+ and Mg^{2+} ions may correlate with the unusual base-pair stacking observed at the CA/TG step (see chapter 2).

Tl^+ 106 is coordinated by O2 of T(8) and O4' of G(9). This mode of coordination has some features in common with the coordination of monovalent cations observed at the central ApT step in the A-tract minor groove of CGCGAATTCGCG. That site is referred to as P3 site (Shui, Sines et al. 1998; Woods, McFail-Isom et al. 2000). Various monovalent cations have been reported to occupy P3 site, including Tl^+ (Howerton, Sines et al. 2001; Moulaei, Maehigashi et al. 2005) Rb^+ (Tereshko, Minasov et al. 1999), Na^+ (Shui, McFail-Isom et al. 1998), K^+ (Shui, Sines et al. 1998) and Cs^+ (Woods, McFail-Isom et al. 2000). Additionally, Tl^+ ions identified in the complex of N-terminal domain of P22 c2 repressor protein and operator sequence of CATTTAAGATATCTTAAATG /CATTTAAGATATCTTAAATG also show similar intra-strand chelation by O2 of thymine and O4' of the following residue (J.D. Watkins and L.D. Williams, unpublished).

However, some of the coordinating ligands of the P3 site of CGCGAATTCGCG are not observed with Tl^+ 106. Differences in position and coordination of Tl^+ 106 in comparison with cation in the P3 site appear to correlate with the minor groove differences between CGCGAATTCGCG and CCAGGCCTGG. The P3 site is coordinated by the second O2 atom of thymine from the opposing strand and the second O4' atom from the opposing strand. The lack of inter-strand coordination with Tl^+ 106 may correlate with a structural difference observed between CGCGAATTCGCG and CCAGGCCTGG. The minor groove is narrow in CGCGAATTCGCG but not in CCAGGCCTGG. A correlation between the interaction of monovalent cations by inter-strand thymine oxygens and the narrow minor groove of A-tract DNA has been previously proposed (Hamelberg, McFail-Isom et al. 2000). Monovalent cation at P3 site of CGCGAATTCGCG is at the center of characteristically narrowed A-tract minor groove, which enables coordination of

monovalent cation by inter-strand thymine oxygens. On the other hand, the corresponding site is occupied by guanine in CCAGGCCTGG, with electropositive amino group (N2) instead of O2 of thymine at the respective position. Tl^+ 106 is thus effectively shifted toward one strand of CCAGGCCTGG, away from N2 of G(14). The width of minor groove of CCAGGCCTGG is wider, which allows inter-strand contacts by O2 of T(8) and N2 of G(14) to first shell water molecules of competing Mg^{2+}.

Conclusion

Local variations observed between Tl^+/Rb^+ positions in this study and previously determined Tl^+ positions in duplex of CGCGAATTCGCG appear to arise from variation in DNA conformation that is sequence dependent. Twelve Tl+ sites and four Rb+ sites are identified in $[d(CCAGGCCTGG)]_2/Tl^+$ and $[d(CCAGGCCTGG)]_2/Rb^+$ respectively. Partially dehydrated monovalent cations are coordinated by oxygen and nitrogen positions of DNA bases. In the major groove, monovalent cations are coordinated with N7 and/or O2 of the adjacent guanine, and additionally an intra-strand N7 of the preceding purine residue (adenine or guanine). The latter was not observed with previously study as there is no purine-purine step in CGCGAATTCGCG with Tl^+ bound guanine at the second position of the step. In minor groove, monovalent cations are coordinated with O2 of thymine and O4' of preceding deoxyribose. Interaction by two inter-strand thymine oxygen atoms (O2) seen in the dodecamer is not seen here, as guanine occupies the corresponding site with electropositive amino group (N2) in CGCGAATTCGCG.

The number of observed Tl^+ sites in $[d(CCAGGCCTGG)]_2/Tl^+$ notably exceeds the number of observed Rb^+ sites in $[d(CCAGGCCTGG)]_2/Rb^+$. Difference in movalent cation distributions between two are likely to arise from physical properties of the cations used.

Rb^+ scatters less strongly than Tl^+ and provides weaker anomalous contribution. All identified Tl^+ sites except for Tl^+ 101 site are located on top of either Rb^+ or water molecules in refined models of $[d(CCAGGCCTGG)]_2/Rb^+$ and $[d(CCAGGCCTGG)]_2/Na^+$ from previous study. This is likely to indicate that identified Tl+ sites may be occupied by either Rb^+ or Na^+ in these x-ray derived models of CCAGGCCTGG, yet unidentified due to difference in experimental signal/noise ($Tl^+ > Rb^+ > Na^+$). At Tl^+ 101 site, high occupancy of competing Mg^{2+} (90%) appears to influence the position of Rb^+/Na^+ and/or interpretation, as less than 10% occupied Rb^+ (3 *Coulombs*) and Na^+ (1 *Coulombs*) may not be distinguishable from nearly overlapping first shell water of Mg^{2+} complex.

Additionally many notable features observed in the refined model of $[d(CCAGGCCTGG)]_2/Na^+$ are verified, such as alternate states observed in both DNA and Mg^{2+} coordinates. Providing ultra-high resolution data of particularly $[d(CCAGGCCTGG)]_2/Rb^+$ with higher resolution and nearly 50% more reflections compared to that of Na^+ form of CCAGGCCTGG, the correlations observed between alternate states of DNA and Mg^{2+} appear to be confirmed.

Database accession number

The structure of $[d(CCAGGCCTGG)]_2/Tl^+$ has been assigned PDB ID of 3GGI. The structure of $[d(CCAGGCCTGG)]_2/Rb^+$ has been assigned PDB ID of 3GGK.

CHAPTER 4

CRYSTAL STRUCTURE OF PROFLAVINE IN COMPLEX WITH A DNA HEXAMER DUPLEX

Abstract

Here, we report the 1.1Å resolution x-ray crystal structure of an intercalative complex between $[d(CGATCG)]_2$ and proflavine. Two proflavine molecules are observed to intercalate at the CpG steps at the both terminals of DNA duplex. Proflavine orients itself so that its terminal amines are toward the major groove. Proflavine molecules are in close van der Waals contacts with the adjacent C/G and G/C base pairs, positioning themselves in the pockets created by these unwound steps. The mode of intercalation by proflavine molecules is observed to be largely through stacking interaction. High planarity and lack of any out-of-plane segment of proflavine enable orientation of the molecule to engage in extensive stacking contacts, therefore anchoring itself at the intercalating site. Helical parameters of the current structure reveal that intercalation causes DNA to unwind and lengthen dramatically at the intercalation sites. Intercalation by proflavine causes DNA to unwind by approximately 13° and lengthen by approximately 3.3Å with respect to standard B-DNA at each site. These changes in helical parameters are found to be similar to the published intercalative complex of DNA with another high planarity intercalator, ellipticine.

Introduction

Noncovalent interactions direct assembly and synthesis of life's polymers. The structural and energetic rules governing process such as nucleic acid replication and protein folding, along with molecular recognition of these biopolymers by various ligands, results from noncovalent interactions. These interactions are generally classified as hydrogen bonding, charge-charge, dipole-dipole, dipole-induced dipole, and dispersion. In a binding or folding reaction, noncovalent interactions generally contribute to changes in enthalpy. Entropic effects arise from ion and water release, changes in conformational or translational degrees of freedom, and a subset of phenomena referred to collectively as the hydrophobic effect. Our understanding of these fundamental aspects of macromolecular folding and recognition is currently not sufficient to allow prediction of many non-canonical nucleic acid structures and drug-nucleic acid interactions.

The function of numerous synthetic and natural products that bind to DNA is determined by mode of binding (Suh and Chaires 1995), sequence preference (Muller and Crothers 1975), conformational preference, and alteration of DNA conformation (Williams, Egli et al. 1992). Each of these parameters is determined by a subtle interplay of molecular interactions. DNA intercalation, as first proposed by Lerman (Lerman 1961; Lerman 1964), is analogous to insertion of a 'false coin' (an intercalator) into a roll of pennies (base pairs). To intercalate in DNA, a planar molecule slides between two base pairs without breaking Watson-Crick hydrogen bonds. Intercalation extends the DNA helix and shifts the base pairs flanking the binding site in opposite directions along the helical axis. Lerman's proposal that planar molecules can intercalate in DNA has been validated by a wealth of physical-

chemical data (Berman and Young 1981; Waring 1981; Neidle and Abraham 1985; Neidle, Pearl et al. 1987; Denny 1989).

Intercalated compounds typically interact with adjacent base pairs primarily through a complex array of molecular interactions (Chaires, Satyanarayana et al. 1996; Chaires 1997). It has been proposed that polarizable compounds intercalate preferentially at sites containing G/C base pairs since their dipole moment is greater than that of their A/T counterparts (Muller and Crothers 1975). Relationships between intercalation and helical unwinding remain unclear. The absence of a plausible molecular basis for helical unwinding derives primarily from a lack of understanding of helical twist in general. It is probably impossible to decipher the causes of intercalator-induced DNA unwinding until the molecular basis for the helical twist of unperturbed DNA is understood. The structure of a DNA•proflavin complex described here allows us to assess the interactions that govern binding orientation and sequence selectivity, and to test and extend previous hypotheses on relationships between intercalator structure and DNA distortion (Williams, Egli et al. 1992).

Materials and Methods

Crystallization and Data Collection

Reverse-phase HPLC purified d[CGATCG] was annealed slowly from 90°C in a water bath upon reconstitution to the desired concentration in nano-pure water prior to the crystallization experiments. Crystals of $[d(CGATCG)]_2 \cdot$ Proflavine complex were grown by vapor diffusion method, either from sitting or hanging drop, initially containing 0.7mM of the ammonia salt of reverse-phase HPLC purified d(CGATCG), 0.7mM of proflavine hemisulfate (Sigma) and crystallization solution. A wide variety of crystallization conditions produce crystals with similar diffraction quality; either 10mM Sodium cacodylate buffer

(pH 5.5) or 12.5mM HEPES buffer (pH 7.0), Magnesium chloride ranging between 0 - 5mM, Cobalthexammine chloride ranging between 0.25 - 0.5mM, and one of the PEG400, MPD or MPD/PEG1000 mix ranging between 6 - 10% in the crystallization drop as the primary precipitant. These drops were equilibrated against the corresponding precipitant with concentrations ranging from 25 to 40%. The structure presented here is based on the crystal grown from a sitting drop initially containing d(CGATCG) and the drug at above mentioned concentration, 12.5mM HEPES (pH 7.0), 0.25mM Cobalthexammine chloride and 9% MPD/PEG1000 mix (3:1), equilibrated against the reservoir solution containing a mix of MPD and PEG1000 at the concentration 25% and 7.5% respectively. In most cases, yellow hexagonal crystals grew to 0.2-0.3mm at its longest edge, typically in 5 - 6 days at 20°C. The data was collected on the Quantum 315 detector (ADSC) at beamline X25 of the National Synchrotron Light Source (Brookhaven, NY). A total of 92739 reflections extending to 1.1 Å were collected from a single crystal at 120K for 360° of reciprocal space data using the wavelength of 1.0 Å. The data was processed and scaled to 10189 unique reflections using the HKL2000 (Otwinowski and Minor 1997). Unit cell dimensions are a=b=24.56 Å, c=78.28Å, α=β=90°, γ=120° in space group $P6_5$. The statistics of diffraction data are shown in table 4.1.

Refinement

A starting model consisting of DNA coordinates of a published $[d(CGATCG)]_2$ · ellipticine complex [PDB entry 1Z3F (Canals, Purciolas et al. 2005)] was used for phase solution through molecular replacement, using the program *CNS version 1.1* (Brunger, Adams et al. 1998) along with parameters of Berman and co-workers (Berman, Zardecki et al. 1998). The improved starting model that consists of only DNA from the program

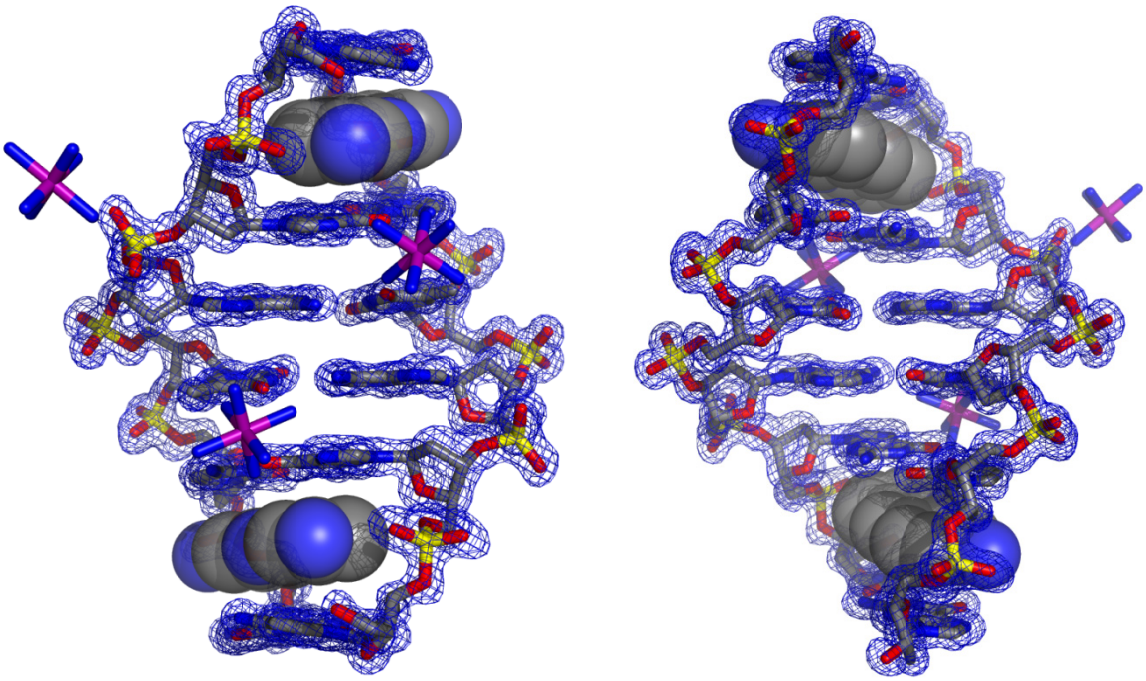

Figure 4.1: Overall view of the 1.1Å crystal structure of [d(CGATCG)]$_2$/Proflavine$_2$ complex. Viewing into the major groove (Left) and the minor groove (Right). DNA strands are shown in stick models, colored with CPK coloring scheme. Sum electron density ($2F_o$-F_c) is shown in blue net around DNA, contoured at 1.0σ. Both proflavine molecules are shown in sphere. Three partially occupied cobalt hexammine complexes are shown in stick model with their metal center colored purple. Water molecules are omitted for clarity. Unless otherwise noted, all figures in this chapter are rendered with Pymol v1.1 (DeLano).

CNS 1.1 through a several round of simulated annealing and refinement was directly fed to the program *REFMAC5* (Murshudov, Vagin et al. 1997). A care was taken in terms of preserving the flagged intensities for *Free-R test* between these two programs. The original intensity file (.sca) was converted to MTZ using the program *sca2mtz* first, and then this MTZ output was processed by the program *Mtz2various* to be used in *CNS*. Both utilities are available in *CCP4 program suit* (Bailey 1994).The same MTZ file was used for the refinement process by the program *REFMAC5*. The model for proflavine molecules were assigned by inspecting the $2F_o$-F_c and the appearance of corresponding F_o-F_c difference Fouriers with the program *Coot* (Emsley and Cowtan 2004), using the ideal coordinates from *HIC-Up server* (Kleywegt and Jones 1998). Three partially occupied cobalthexammine

molecules were also added to the final model, where six peaks of ammine moieties surrounds a large $2F_o$-F_c peak (8-16σ, depending on occupancy) and the corresponding F_o-F_c of cobalt metal center in the octahedral geometry. Water molecules were also added to peaks of corresponding $2F_o$-F_c and F_o-F_c peaks visualized in the program *Coot* (Emsley and Cowtan 2004). Hydrogen atoms have been added in the riding positions during the refinement process (with no output to the final PDB), and all atoms other than water molecules in the model were refined anisotropically. Refinement process was continued until the convergence of refinement statistics was reached, while R-free and thermal ellipsoids were carefully monitored to avoid over refinement. The refinement statistics are summarized in table 4.1. Thermal ellipsoids (figure 4.2) were computed with *ORTEP-3 for windows* (Farrugia 1997)**.** Superimpositions with other published structures were performed using the McLachlan algorithm (McLachlan 1982) as implemented in the program *ProFit* (Martin, A.C.R., `http://www.bioinf.org.uk/software/profit/`). The helical parameters of the DNA was analyzed with the program *CURVES* (Lavery and Sklenar 1988).

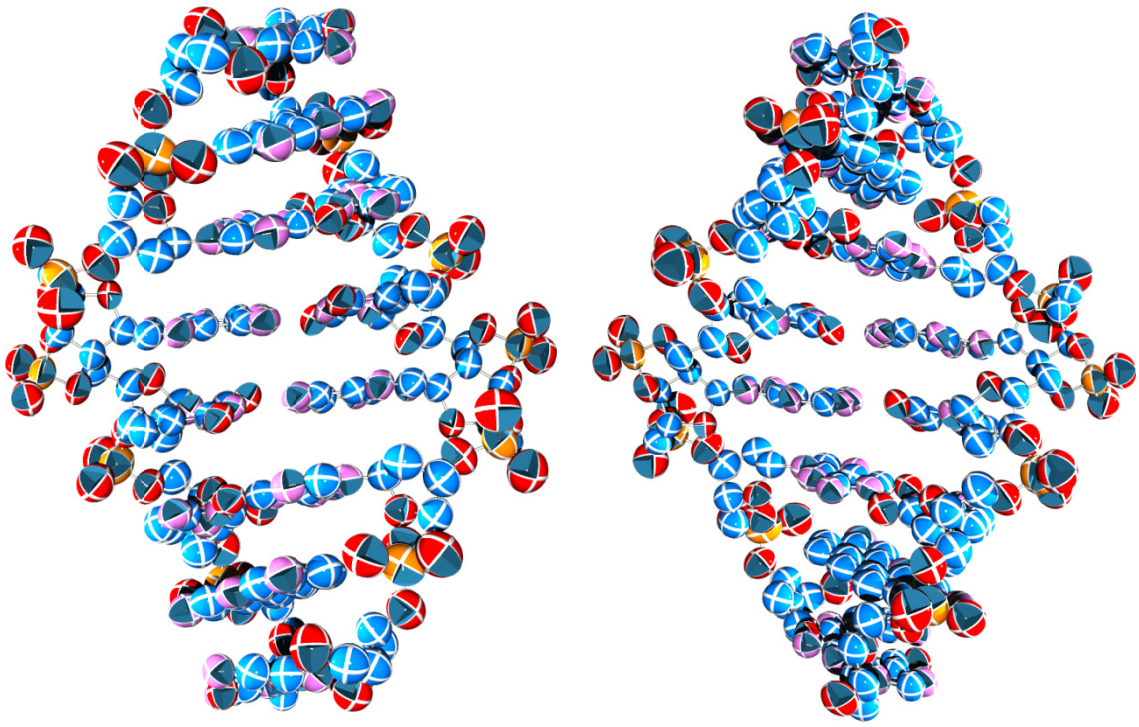

Figure 4.2: Thermal ellipsoids of the 1.1Å crystal structure of [d(CGATCG)]$_2$/Proflavine$_2$ complex. Viewing into the major groove (left) and the minor groove (right). Both plots are shown at 50% probability. Average isotropic atomic displacement (B-value, overall) is determined to be 15.45Å^2.The plots were calculated with Ortep-3 for windows (Farrugia, 1997) and rendered with Povray 3.6. Cobalt hexammine complexes are omitted for clarity. In this structure, water molecules were not refined anisotropically.

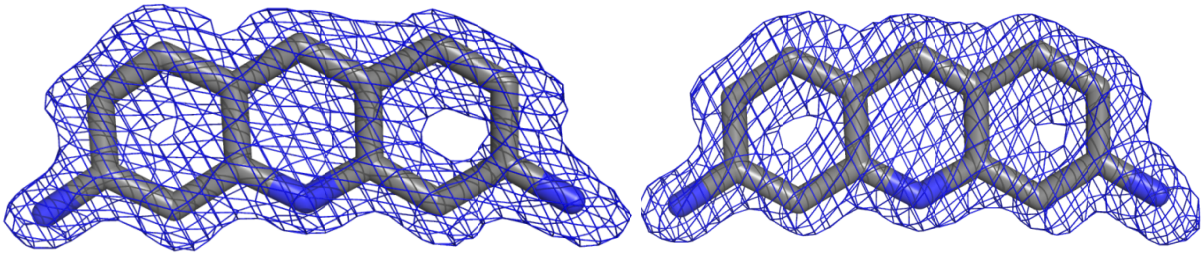

Figure 4.3: Sum electron density (2Fo-Fc) surrounding proflavine molecules. Proflavine molecules are shown in stick models, colored with CPK coloring scheme. Sum electron density ($2F_o$-F_c) is shown in blue net around proflavine, contoured at 1.0σ. (Left) Proflavine 21 (Right) Proflavine 22. Notice the clean and continuous surrounding electron density, clearly indicating the orientation of the terminal amine (toward the major groove)

Table 4.1: Crystallographic and Refinement Statistics

Data Collection	
Source wavelength (Å)	0.99997
Space group	$P6_5$
Unit cell (Å)	a = b = 24.563
	c = 78.276
Resolution range (Å)	21 – 1.12
Number of reflections	92542
Number of unique reflections	9904
Completeness (%)[a]	96.6 (73.4)
Average I/σ (I)	44.8 (1.75)
R_{merge} (%)[b]	5.2 (33.4)
Refinement statistics	
DNA (asymmetric unit)	$d[CGATCG]_2$
Number of DNA atoms	241
Number of water Molecules[c]	37
Number of Cobalthexammine ions[c]	3
R_{work} (%)	18.4
R_{free} (%)	21.9 (1096 reflections)
RMS deviation of bonds from ideal (Å)	0.010
RMS deviation of angles from ideal (°)	1.732
Average isotropic B value	15.45
PDB ID code	3FT6

[a] The values in parentheses refers to the highest resolution shells
[b] $R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I = observed intensities and $\langle I \rangle$ = mean intensity obtained from multiple observations of symmetry-related reflections and rejections.
[c] Includes partially occupied species. Disordered species with same residue number are counted as single molecules.
[d] $R_{work} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$, where Fo and Fc are the observed and calculated structure factors, respectively. Reflections flagged for Free R test (7.8%) are excluded from the calculation. R factor for all reflections is determined to be 18.81%.
[e] R_{free}= defined by Brünger (Brunger 1992)

Results

A total of 9904 reflections in the resolution range 20.5 - 1.12 Å were used in the refinement, including randomly selected 1 096 reflections (11.1%) for cross validation. There is no exclusion of the reflections by sigma-cut. The final R- factor is determined to be 18.81% for all data. The crystallographic asymmetric unit consists of a duplex of the hexamer, $d(CGATCG)_2$ and two intercalated proflavine molecules, assembling a biological unit. As customary, bases are number 1-6 in the first strand, 7-12 in the other. Both proflavine molecules are observed to intercalate at the CpG steps at the both terminals of the duplex. Proflavine orients itself so that its terminal amines are toward the major groove. All atoms of both proflavine molecules fit cleanly into the $2F_o$-F_c electron density and are well ordered (figure 4.3). The $2F_o$-F_c map around DNA is also clean and continuous (figure 4.1). Three partially occupied cobalt hexammine molecules (NCO 31 – 33) are also observed per asymmetric unit, with estimated occupancies ranging from 50 to 70%. These values of partial occupancies correlate with the low availability of the cobalt hexammine molecules in the crystallization setup with respect to the available DNA/drug complex (0.25mM vs. 0.35mM $[d(CGATCG)]_2 \cdot Proflavine_2$). Two of these ions are localized in the major groove in proximity to guanine bases; 3Å from N7 and O6 of G (11) to N1 of NCO (32), 3Å from N7 of G(2) to N1 and N2 of NCO (33), and 3.17Å from O6 of G (2) to N5 of NCO (33). NCO (32) also makes a contact with the terminal amine of one of the proflavine (at the opposite side from the backbone contact) at a distance of 3.47Å (see below). While 2Fo-Fc density show most octahedron of ammonia clearly coordinating to each cobalt center with peaks of >2σ, a few of the coordinated

ammonia at lower occupied sites (50%) were observed with smaller peaks of approximately 0.8σ.

Global helix structure

Helical parameters of the current structure reveal that intercalation causes DNA to unwind and lengthen dramatically at the intercalation sites. The DNA helical twist and rise at the first intercalation site by proflavine between C(1)/G(12) and G(2)/C(11) base pairs is 23.6° and 6.7Å. The second intercalation site between C(5)/G(8) and G(6)/C(7) pairs show the similar corresponding values of 22.9° and 6.7Å. These values are significantly different from average B-DNA values of ~36° (helical twist) and ~3.4Å (rise). Whereas non-intercalative steps show comparable values of 34-36° of helical twist and 3.1-3.6Å of rise, indicating that the rest of the DNA is almost in canonical B-form. Intercalation by proflavine therefore causes DNA to unwound approximately 13° and lengthen approximately 3.3Å with respect to standard B-DNA at each site. These changes in helical parameters are found to be similar to the published $[d(CGATCG)]_2$·ellipticine complex [PDB ID 1Z3F (Canals, Purciolas et al. 2005)], also indicated by rmsd of 0.168 (all DNA atoms excluding terminal oxygen atoms) when superimposed (figure 4.4).

This major conformational change also affects the geometry of DNA backbone including sugar puckering. The mixed sugar puckering at the immediate site of proflavine intercalation is apparent, especially at the bases toward both terminals of the helix. At the first site of proflavine intercalation, the CpG step, C(1) sugar is found to adapt C3'-*endo* sugar pucker, while G(2) conserves C2'-*endo* typically seen in B-DNA. Same is true for the opposing strand, as G(12) is observed to be in O4'-*endo*, whereas C(11) shows C1'-*exo* that is in neighboring of the C2'-endo domain (Altona and Sundaral.M 1972). The

same trend continues with the second proflavine intercalation site. Sugar puckering s found in the both terminal bases, G(6)/C(7) are the most deviated from the normal puckering mode (O4'-*endo* and C3'-*endo* respectively). Whereas C(5)/G(8) bases are found to adopt sugar pucker in or adjacent to the C2'-*endo* domain (C1'-*exo* and C2'-*endo*, respectively).

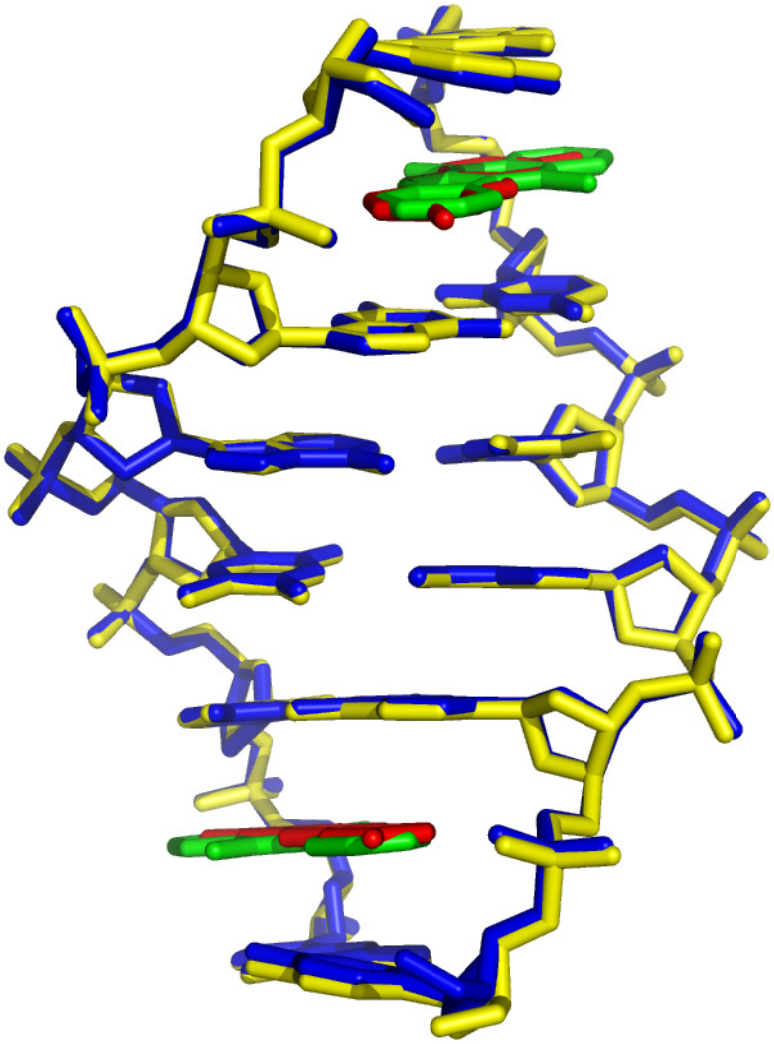

Figure 4.4: [d(CGATCG)]2 · elipticine complex (PDB entry: 1Z3F) superimposed with the current structure of [d(CGATCG)]$_2$ · proflavine complex. The proflavine and elipticine molecules are shown in green and red, respectively. DNA coordinates reported in 1Z3F is shown in yellow, DNA coordinates in the current structure is shown in blue. R.m.s. deviation for all atoms in DNA (excluding terminal oxygens) is 0.168Å. Superimpositions was performed using the McLachlan algorithm (McLachlan 1982) as implemented in the program ProFit (Martin, A.C.R., http://www.bioinf.org.uk/software/profit/).

Proflavine contacts

The mode of intercalation observed in [d(CGATCG)$_2$·proflavine$_2$ complex here is essentially in accord with previously described crystal structure of dinucleotide ·proflavine complex [NDB ID ddb009 (Berman, Stallings et al. 1979; Shieh, Berman et al. 1980)] (figure 4.5). Proflavine molecules are in close van der Waals contacts with the adjacent C/G and G/C base pairs, positioning themselves in the pockets created by these unwound steps (figure 4.6). The mode of intercalation by proflavine molecules is observed to be largely through stacking interaction. High planarity and lack of any out-of-plane segment of proflavine enable orientation of the molecule to engage in extensive stacking contacts, therefore anchoring itself at the intercalating site.

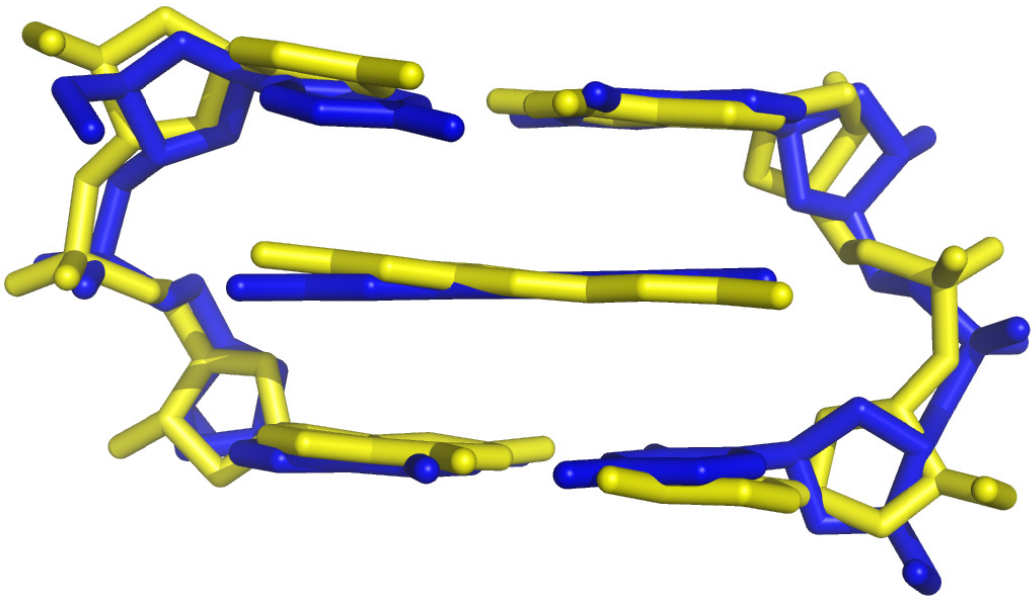

Figure 4.5: Dinucleotide [d(CG)]2 ·proflavine complex (NDB entry:ddb009) superimposed with the corresponding region of the current structure of [d(CGATCG)]$_2$ · proflavine complex. The non-intercalating second proflavine reported in ddb009 is omitted for clarity. Current structure is shown in blue, while ddb009 is shown in yellow. R.m.s. deviation for all atoms in DNA (excluding terminal oxygen atoms) is 0.813Å. Superimpositions was performed using the McLachlan algorithm (McLachlan 1982) as implemented in the program ProFit (Martin, A.C.R., http://www.bioinf.org.uk/software/profit/).

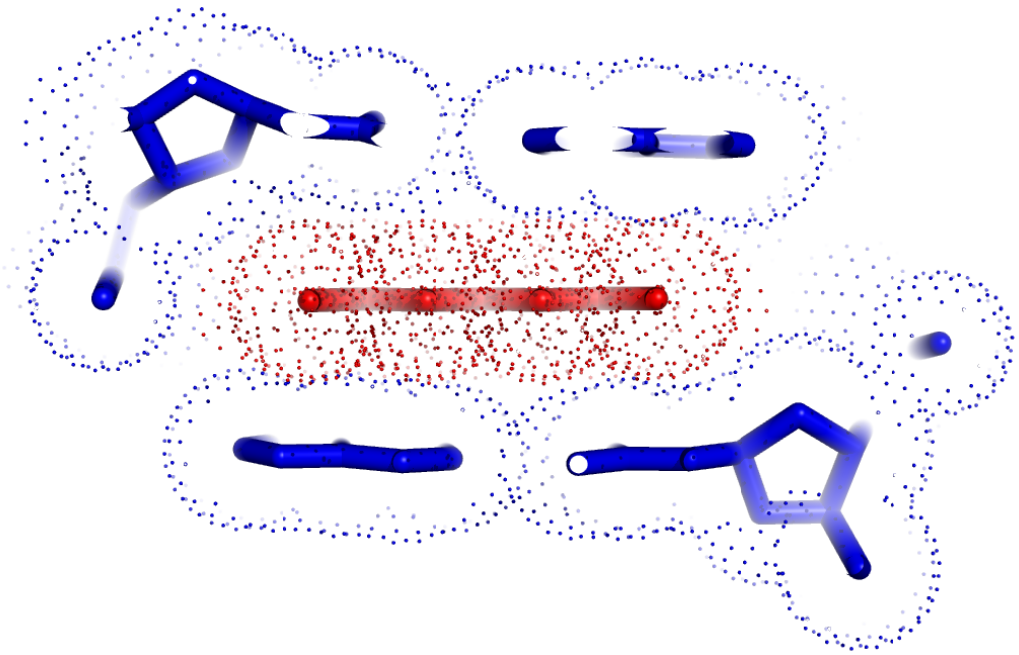

Figure 4.6: A schematic representation at the intercalating site of the structure of $[d(CGATCG)]_2$ · proflavine complex, CG/GC pairs and proflavine 21 are shown. Viewing into the major groove at the intercalating site. DNA base pairs and proflavine molecule are shown in stick model, colored blue and red, respectively. Each atom is surrounded with dots indicating VDW radius (1.2Å) to emphasize encapsulating pocket created by the terminal base pairs to accommodate the intercalator.

There is no direct H-bond interaction observed between proflavine molecules and DNA in the asymmetric unit. However, several water mediated interactions with DNA are observed (figure 4.7, 4.8). Both proflavine molecules (PRL 21 and 22) share a common water-mediated H-bond interaction between positively charged center nitrogen (N10) to N4 of the terminal cytosine residue C(1) or C(7) through HOH 108 or 113, respectively. The same water-mediated H-bond interaction is also seen with cross-strand N4 of another immediate cytosine residue at PRL 22 intercalating site, C(5). Both PRL 21 and 22 form water mediated contacts to the adjacent backbone oxygen atoms (O5' and phosphate oxygen through HOH 110 or 111) through one side of terminal amine groups. In the second proflavine molecule (PRL22), this is altered on the other side, as its

terminal amine forms H-bond interaction with one of the coordinated ammine groups of a neighboring cobalt-heaxammine complex 32 in the major groove. The same ion complex is held between following the GpA step, forming H-bond interactions with both N7 and O6 of the adjacent G(8) and O4 of T(4) from opposing strand, however no interaction is seen with the adjacent phosphate group. This ion-mediated contact is slightly altered with the first proflavine (PRL21), as the closest distance between its terminal amine and one of the coordinated ammine moieties of the neighboring Co^{2+} complex 33 is found to be 3.6Å. Although, the location of Co(33) is similar to that of Co(32), but its orientation is slightly deviated with respect to neighboring bases. Co (33) interact with the same atoms of corresponding neighboring bases, but with three ammine moieties, whereas Co (32) does the same with two.

The $[d(CGATCG)]_2$·proflavine complex is tightly packed in the crystal through hexagonal spacegroup $P6_5$, allowing only 37 water molecules (including partially occupied molecules) per asymmetric unit (figure 4.9). Each end of asymmetric unit invades the minor groove of another asymmetric unit at the ApT step, with direct contact between N1 of terminal guanine and O4' of symmetry related A(3) at the distance of 3.0Å. The helices are also held in place by H-bond interaction through coordinated ammine moiety of Co (31) between O4' of terminal cytosine and O1P of A(3). This unusual pattern of end-to-side interactions is repeated through the crystal lattice, so that every asymmetric unit is both invaded in the minor groove and invader at each end. Different units of these pseudo-continuous helices meet laterally as phosphate groups of one asymmetric unit are in proximity to the major groove of another asymmetric unit. The phosphate oxygen atoms of T(4) and C(5) of one helix are held in place by the major

groove bound cobalt hexammine complex of another helix, either by Co (32) or Co (33) mentioned above. As a result, one terminal amine of both proflavine molecules shows lattice interaction with a phosphate oxygen atom of a symmetry related unit.

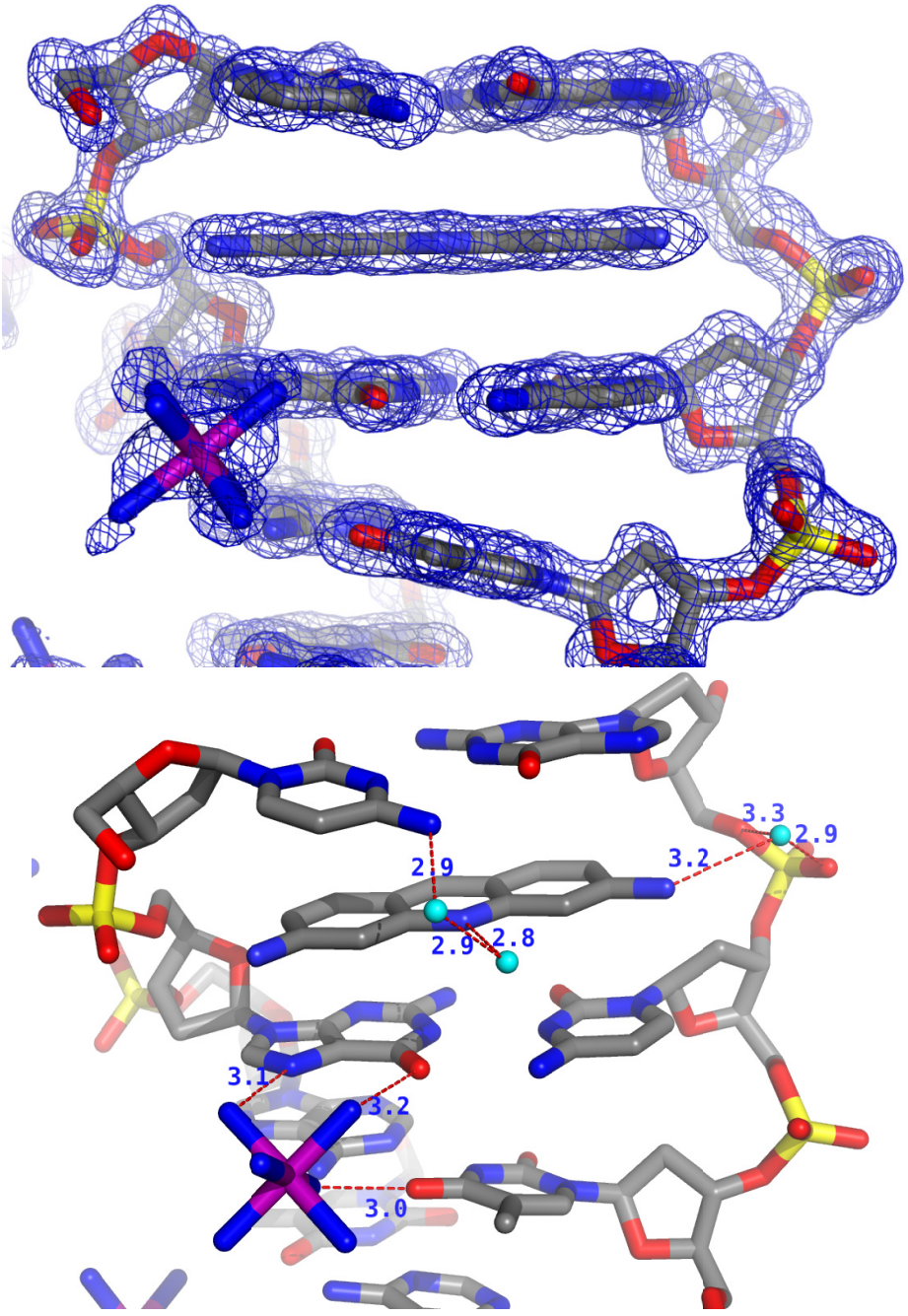

Figure 4.7: Detailed view at proflavine 21 intercalating site of [d(CGATCG)]$_2$/Proflavine$_2$ complex, at the C(1)pG(2) step. In both figures, DNA, proflavine and cobalt hexammine molecules are shown in stick, colored according to CPK coloring scheme. Cobalt metal center is colored purple. (Left) Sum electron density ($2F_o$-F_c) is shown in blue net around DNA, proflavine (21) and cobalt hexammine (33), contoured at 1.0σ. (Right) Contacts between proflavine, cobalt hexammine and DNA with less than 3.5Å are shown. Non contacting water molecules are omitted for clarity. One terminal amine group of proflavine 21 makes water mediated contact with adjacent phosphate group. Notice the lack of cobalt hexammine mediated contact on the other side of proflavine (see figure 4.8). For reference, contact distance between center N atom of proflavine and N4 of terminal cytosine, C(1) is 3.3Å. Terminal amine to N7 of G(12) is 3.4Å. Although, these contacts are with the geometry that is not appropriate for hydrogen bonding.

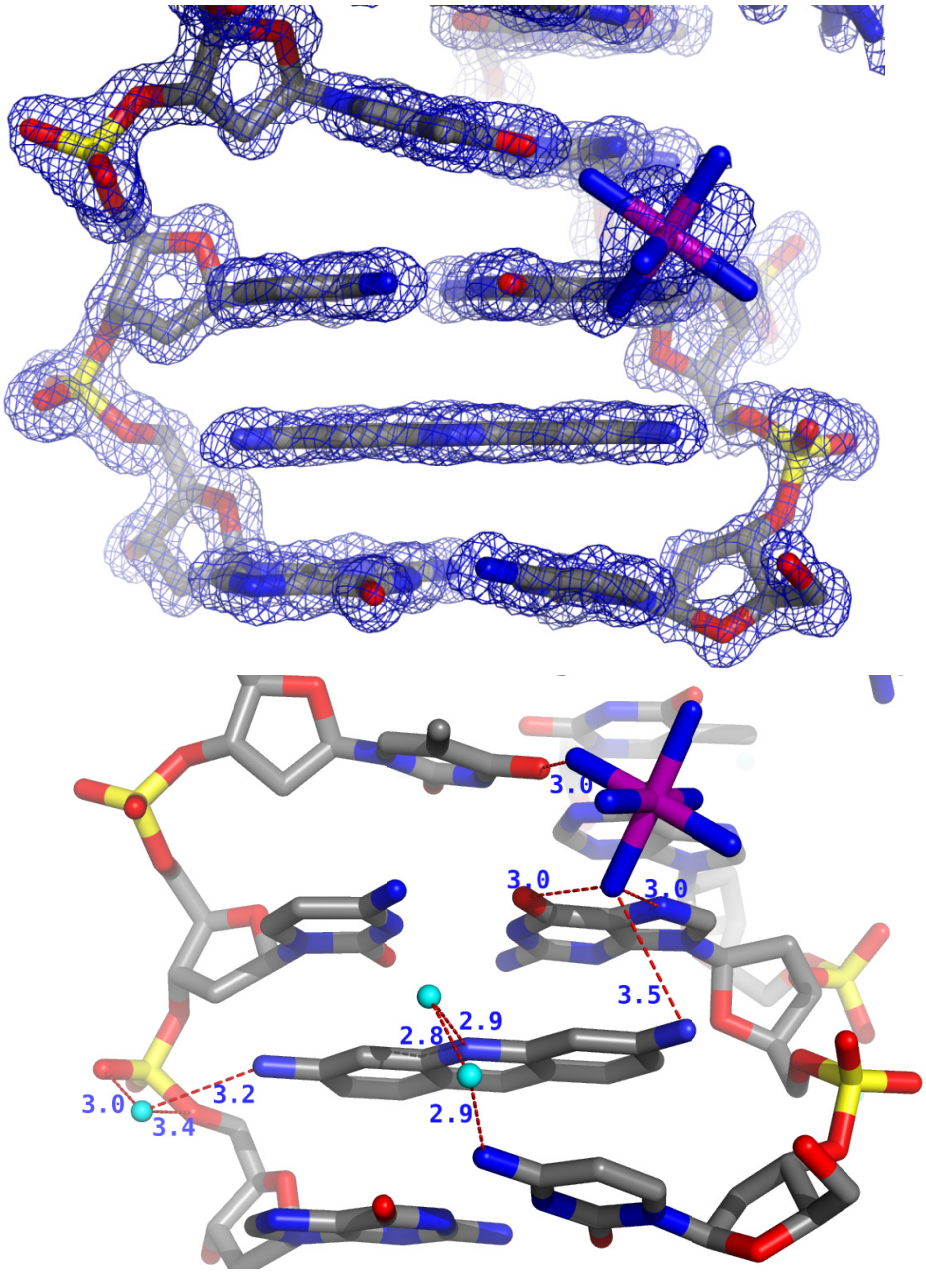

Figure 4.8: Detailed view at proflavine 22 intercalating site of [d(CGATCG)]$_2$/Proflavine$_2$ complex, at the C(5)pG(6) step. In both figures, DNA, proflavine and cobalt hexammine molecules are shown in stick, colored according to CPK coloring scheme. Cobalt metal center is colored purple. (Left) Sum electron density ($2F_o$-F_c) is shown in blue net around DNA, proflavine (22) and cobalt hexammine (32), contoured at 1.0σ. Estimated occupancy for Co^{2+} 32 complex is 50%. Some ammine moiety are observed to have weak sum density peaks at approximately 0.8σ. (Right) Contacts between proflavine, cobalt hexammine and DNA with less than 3.5Å are shown. Non contacting water molecules are omitted for clarity. One terminal amine group of proflavine 22 makes water mediated contact with adjacent phosphate group. Notice the contact with neighboring cobalt hexammine 32. Alignment of this Co^{2+} complex slightly differs from that of Co^{2+} 33, also indicated by the altered H-bond interactions with adjacent bases (see figure 4.7). For reference, contact distance between center N atom of proflavine and N4 of terminal cytosine, C(1) is 3.3Å. Terminal amine to N7 of G(12) is 3.4Å. Although, these contacts are with the geometry that is not appropriate for hydrogen bonding.

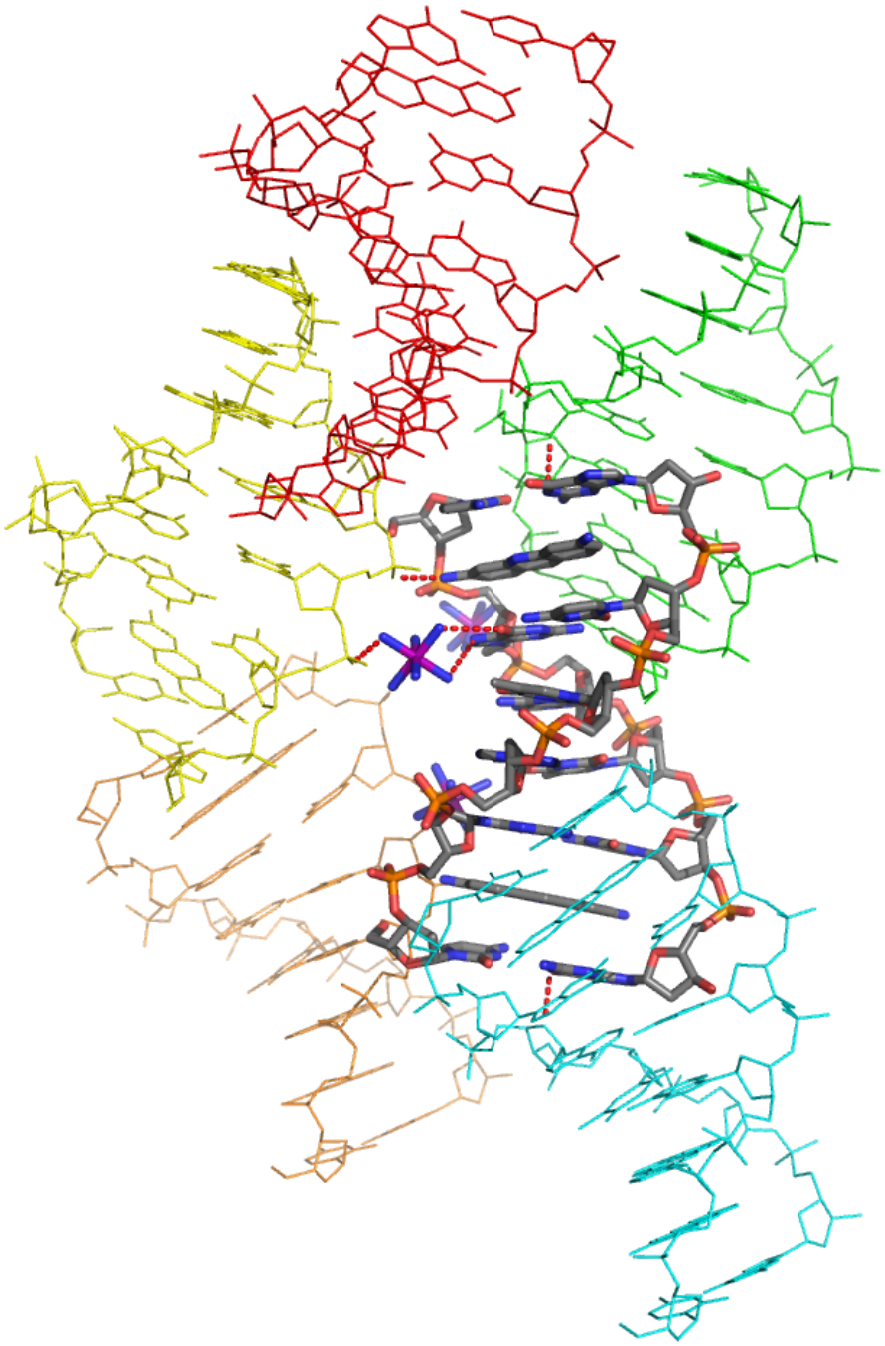

Figure 4.9: Crystal packing of the [d(CGATCG)]$_2$ · Proflavine complex. Five asymmetric units are shown colored differently, around center reference asymmetric unit color according to CPK. Lattice interactions between asymmetric units between reference unit and adjacent unit, as well as cobalt hexammine mediated are shown (<3.5Å). In crystal space group P65, each unit is both invaded and invader at each end, making direct H-bond interaction between N1 of terminal guanine and O4' of A(3). Lateral packing of the crystal causes cobalt complex-mediated interactions between units (shown at the center), and one terminal amine of proflavine makes direct lattice interaction with phosphate oxygen atom of the adjacent asymmetric unit.

Discussion

Orientation and sequence selectivity of proflavine

The structure of $[d(CGATCG)]_2$·proflavine complex reported here is in accord with the sequence selectivity of proflavine to G/C base pairs (Muller and Crothers 1975), as both proflavine molecules are observed to intercalate at the both terminal CpG steps. No proflavine intercalation is observed at the other dinucleotide steps present in the duplex, the GpA, ApT and TpC steps. As there is no direct H-bond interaction observed between proflavine and immediate bases, the mode of intercalation by proflavine molecules appears to be largely through stacking interaction. Both proflavine molecules appear to have the most overlap with the six-membered rings of the flanking guanine and cytosine residues (figure 4.10). Proflavine molecules are also in close van der Waals contacts with the adjacent C/G and G/C base pairs, positioning themselves in the pockets created by these unwound steps.

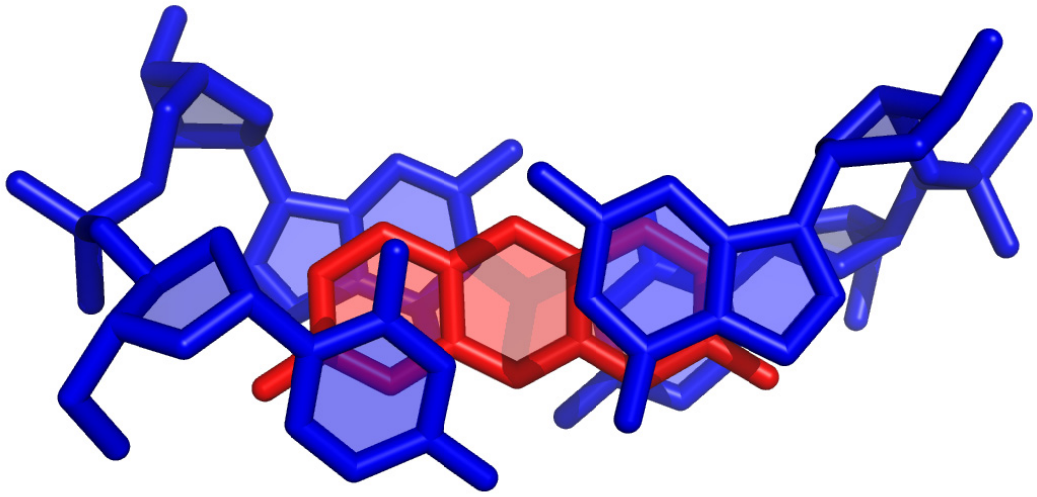

Figure 4.10: Axial view of the $[d(CGATCG)]_2$ · Proflavine complex. Proflavine 21 and the C(1)pG(2) steps (on left) are shown. Terminal base pairs are colored blue, Proflavine molecule is colored red. Notice the stacking between six-membered rings of the flanking guanine and cytosine residues with proflavine. For contacts with neighboring water and cobalt hexammine, see figure 4.7.

Muller and Crothers suggested that the sequence specificity of proflavine to G/C base pairs originate from its polarizability matches with the dipole moment of G/C that is larger than that of their A/T counterparts (Muller and Crothers 1975). This appears to be valid with the current observation because both proflavine molecules orient themselves so that their terminal amines are toward the major groove. This orientation is expected to be sterically favored. If we assume that proflavine is readily polarized in the direction from the major to the minor groove ($\delta^+ \rightarrow \delta^-$) due to the charged nitrogen now facing the major groove, this polarization of the intercalated proflavine provides a favorable polar interaction with the dipole moment of adjacent guanine residues. Both terminal amine groups of each proflavine molecules display close distance of less than 3.5Å to the N7 of adjacent guanine residues. Although the geometry is not appropriate for H-bond interactions, the long-range electrostatic interaction may also correlate with the stability of intercalated proflavine molecules. The major groove side of guanine base has been previously characterized as the electrostatistically most favorable loci (Pullman and Pullman 1981). The site is recognized as a common binding site for both divalent and monovalent cations, as described in previous chapters.

Interactions of cobalt hexammine complexes

At the TpC step in each strand, large degree of tilt (approx. +/- 9°) is observed, which appears to be induced by the large negative propeller twist (approx -15°) at each central A/T base pair. As a result, each strand at the TpC step shows slightly unwound local helical twist of ~32°, which is also indicated by the close distance of ~6.1Å between adjacent phosphate groups. While it is possible that these alterations at two central base pairs are the results of maximizing stacking interactions between intercalators and the

both terminal CpG steps, the neighboring major-groove bound cobalt hexammine molecules appear to stabilize such conformations through additional interaction with inter-strand O4 of thymine residue. In crystal structure of $[d(CGATCG)]_2$/ellipticine complex (Canals, Purciolas et al. 2005), another similarly sized planer intercalator, the TpC step of each strand shows similar tilt (approx. +/-11°), and nearly identical values of propeller twist and slight local unwound at the central A/T base pairs. Consequently, the ellipticine complex was also crystallized in the presence of cobalt hexammine molecules. Although the authors note that the identification of the ammine moieties were impossible due to the lower resolution of the structure, similarly positioned cobalt ions indicate that the cobalt hexammine complexes are equally in effect of stabilizing twisted thymine residues as seen here in hexamer/proflavine complex.

Unwinding of DNA

Both DNA dinucleotide [NDB ID ddb009 (Berman, Stallings et al. 1979; Shieh, Berman et al. 1980)] and DNA hexamer (this work) in complex with proflavine molecule have similar helical geometry at the intercalation site. Sugar puckering is affected and *C3'-endo* is observed at 5' side of each intercalation site. Intercalation by proflavine molecule causes DNA to dramatically lengthen by 3.1-3.3Å. Phosphate at the intercalated CpG step is observed to be always in B_I helical subtype as torsion angle ε is observed to be in *trans* and ζ in (-) *gauche* conformation in both dinucleotide and hexamer cases. However, the apparent difference between proflavine intercalated DNA dinucleotide and hexamer is the extent of DNA unwinding.

In the average B-form of DNA the helical twist is 36°. Each base pair is rotated around the helical axis by this amount relative to the preceding base pair. Intercalation

invariably decreases the helical twist, unwinding the DNA in the vicinity of the binding site to less than 36° per base pair. The helical twist of $[d(CG)]_2$/Proflavine is observed to be 16.7°, which is approximately 19° of unwinding compare to 36° helical twist of an average B-DNA. Whereas the helical twist of $[d(CGATCG)]_2$ at each intercalation site is 23.5°, therefore causing DNA to unwind by approximately 12.5°. There is thus a net difference of >7° of unwinding caused by proflavine intercalation between these two set of host DNA. The apparent net difference appears to be caused by the degree of freedom given to the dinucleotide as it lacks the subsequent or the preceding base steps next to the intercalation site compare to the DNA hexamer in this study. In other words, dinucleotide/proflavine complex is an ultimate case where DNA can alter its conformation to maximize the stacking with intercalator without hindered by intrinsic forces of DNA, such as electrostatic repulsion between adjacent phosphate groups. In the case of the hexamer/proflavine complex, further winding by intercalation may therefore be restricted by such intrinsic forces arising from the existence of subsequent/preceding base pairs.

Williams and Egli proposed a simple model in which DNA helical twist is determined predominantly by a balance of two types of interaction, one that tends to wind the helix opposed by a second that tends to unwind the helix. Electrostatic repulsion between adjacent phosphate groups tends to wind the helix; winding increases the distances separating adjacent phosphate groups. Whereas base-base stacking interactions tend to unwind the helix by increasing the extent of van der Waals contacts between base pairs (Williams, Egli et al. 1992). The proposal was made based on the observations that, although intercalation lengthens DNA, x-ray crystallographic structures of intercalated

DNA demonstrate that the distances between adjacent phosphate groups do not increase. In B-DNA, which displays a helical twist of 36°, the distance between adjacent phosphate groups averages 6.7 Å. In intercalated complexes, B-like distances are maintained between adjacent phosphate groups because unwinding compensates for lengthening. This is as opposed to the common assumption that intercalation-induced DNA unwinding enables the sugar-phosphodiester backbone to span the bound intercalator and still maintain the link between the two flanking base pairs.

In $[d(CGATCG)]_2$·proflavine complex, the average distance between adjacent phosphate groups is 6.7Å at both intercalating sites. This is comparable to the average distance of 6.5Å, between adjacent phosphate groups elsewhere in the complex. Whereas intercalation by proflavine both lengthens and unwinds the DNA by 3.3Å and 13° at each intercalation site. While this invariance supports the importance of electrostatic repulsion by adjacent-phosphate in DNA winding and in DNA conformation in general, it appears that the lowest energy conformation of DNA is achieved by the greatest degree of stacking (i.e., the greatest helical unwinding) allowed by repulsion between adjacent phosphate groups as suggested previously.

Database accession number

The structure of $[d(CGATCG)]_2 \cdot Proflavine_2$ complex has been assigned PDB ID of 3FT6.

APPENDIX A

ESTIMATING OCCUPANCIES WITH SHELX97

Occupancies of the atoms in each conformation within a disordered region can be estimated through visual inspection of the sum densities as well as disappearance/appearance of difference densities surrounding the region. For the structures presented in this theses, the refinement program SHELX97 (ref) was often employed to perform occupancy refinement to estimate the initial assignment of the occupancies. This, in turn, accelerated the whole refinement process with higher accuracy, compare to manual estimation solely based on visual inspection of the electron densities. To indicate two components disordered region in SHELX instruction file (.ins), firstly, "PART" instruction was placed in front of coordinates indicating each region, with the designated number, for example, "PART 1" for the conformation A and "PART 2" for conformation B. The PART instruction can be terminated with "PART 0" placed at the end of coordinates where no more alternate conformations are assigned.

In SHELX instruction files, occupancies for non-disordered atoms with occupancy of 1 (100%) are indicated as 11 because of the Free Variable Concept that the program employs. Second digit "1" indicates that the occupancies for these atoms corresponding to the overall scale factor (osf), which is defined by the first parameter under FVAR instruction. **`FVAR osf[1] free variables`** Thus, to perform a occupancy refinement for the disordered regions indicated by the PART instructions above, the occupancies for all corresponding atoms in each conformation are replaced by 21 (free variable 2) for the conformation A and -21 for the conformation B

within the first disordered region, 31/-31 for the second disordered region (free variable 3), and so forth. Opposite signs of these pre-assigned occupancies instructs software to estimate occupancies with sum of 1 (100%) within the same disordered region. The starting value for the occupancy refinement can be defined by corresponding FVAR parameter. For example, the first disordered region with assigned occupancy 21/-21 (free variable 2), one can assign 0.6 for the second FVAR parameter. This instructs the program to refine occupancies of all atoms in conformation A in the first disordered region starting from 60%, and the occupancies for the corresponding atoms in conformation B starting from 1 – 0.4 = 0.4 (40%).

```
FVAR       0.79918   0.6
.................
RESI 1001   C
PART    1
O5'    4   -0.252111    0.061418    0.369119    21.00000    0.08233
PART    2
O5'    4   -0.243641    0.091332    0.442406   -21.00000    0.12062
PART    0
.................
```

Figure A-1: Sample Shelx instruction for occupancy refinement for the first disordered region, terminal O5' in residue 1. Each conformation is defined by PART instructions, and the occupancy for the atom in each conformation has been replaced by 21/-21. The second parameter in FVAR instruction 0.6 (60%) indicates the starting value for the occupancy calculation for the first conformation indicated by PART 1, occupancy 21, and 40% (1-0.6=0.4) for the second conformation indicated by PART 2 with occupancy -21.

For the occupancy refinement with more than 2 components in the disordered region, all atoms in each component will be assigned a separate free variable designation, for example 31, 41 and 51, and the starting occupancy for each component will also be defined in corresponding parameters in FVAR instruction, third, fourth and fifth. To

define the sum of occupancies within the region to be 1, a linear restraint instruction, SUMP was used. **`SUMP c sigma c1 m1 c2 m2`** Where c = c1*fv(m1) + c2*fv(m2) ··. When SUMP instruction is used for defining the total occupancy among specific free variables, in this case 3, 4 and 5, the term "c" will be 1 defining the total occupancy to be 1. Sigma value of 0.01 is suggested, and was used throughout. Next term c1, c2 and c3 define the absolute ratio between the components specified by the free variables, hence these terms are generally 1 in the case of occupancy refinement. Lastly, m1, m2 and m3 will specify which free variables to be treated by the SUMP instruction, therefore m1=3, m2=4 and m3=5 in this case.

The usage of occupancy refinement is not limited to the region of multiple conformations, and it was also utilized to estimate the occupancies of stand-alone partially occupied ions such as Thallium and Rubidium explained in later chapters. In the case of these partially occupied atoms, the method is essentially the same as above, using FVAR instruction, however there is no need for PART instruction unless the target atom is part of the region taking multiple conformations.

As noted in this thesis, refinement program REFMAC was used for refining each structure presented in this thesis, hence it should be stressed that the occupancies determined through SHLEX were used only as the initial estimates when assigning multiple conformations and these values are no mean finalized occupancies without visual inspection of electron density maps after each subsequent refinement. This is mainly due to the semi-interchangeable nature between refinement programs, where different algorithm and weighing scheme between multiple could possibly impose altered refinement statistics.

```
.................

SUMP 1 0.01 1 3 1 4 1 5

FVAR       0.79918   0.6   0.33   0.33   0.33   .................

.................

RESI 2005   MO6

PART 1

MG     5    0.136560   -0.158312    0.188958    31.00000    0.16541

O1     4    0.126793   -0.242851    0.188238    31.00000    0.17250

O2     4    0.146032   -0.075839    0.184674    31.00000    0.13704

O3     4    0.202478   -0.173965    0.192163    31.00000    0.13070

O4     4    0.071165   -0.147745    0.191443    31.00000    0.06751

O5     4    0.103386   -0.161649    0.120053    31.00000    0.13121

O6     4    0.167168   -0.159027    0.257634    31.00000    0.17402


PART    2

MG     5    0.192989   -0.083347    0.292169    41.00000    0.16300

O1     4    0.142291   -0.107104    0.312379    41.00000    0.06662

O2     4    0.242208   -0.058001    0.269734    41.00000    0.21024

O3     4    0.175765   -0.002185    0.294523    41.00000    0.16363
```

Figure A-2: Sample Shelx instruction for occupancy refinement for the disordered region containing 3 components, residue 2005 MO6 (hexacoordinated magnesium ion). Each conformation is defined by PART instructions, and the occupancy for the atom in each conformation has been replaced by 31, 41 and 51 (free variable 3, 4, and 5). FVAR instruction defines the starting occupancy for each conformation for the occupancy refinement, and SUMP instruction defines the sum of total occupancies of these 3 conformations is 1 (100%).

REFERENCES

Altona, C. and Sundaral.M (1972). "CONFORMATIONAL-ANALYSIS OF SUGAR RING IN NUCLEOSIDES AND NUCLEOTIDES - NEW DESCRIPTION USING CONCEPT OF PSEUDOROTATION." Journal of the American Chemical Society **94**(23): 8205-&.

Badger, J., A. Kapulsky, et al. (1994). "Structure and selectivity of a monovalent cation binding site in cubic insulin crystals." Biophys. J. **66**(2 Pt 1): 286-92.

Badger, J., Y. Li, et al. (1994). "Thallium Counterion Distribution in Cubic Insulin Crystals Determined from Anomalous X-ray Diffraction Data." Proc. Natl. Acad. Sci. U.S.A. **91**(4): 1224-1228.

Baikalov, I., K. Grzeskowiak, et al. (1993). "The crystal structure of the trigonal decamer C-G-A-T-C-G-6meA-T-C-G: a B-DNA helix with 10.6 base-pairs per turn." J. Mol. Biol. **231**(3): 768-784.

Bailey, S. (1994). "THE CCP4 SUITE - PROGRAMS FOR PROTEIN CRYSTALLOGRAPHY." Acta Crystallographica Section D-Biological Crystallography **50**: 760-763.

Basu, S., R. P. Rambo, et al. (1998). "A Specific Monovalent Metal Ion Integral to the AA Platform of the RNA Tetraloop Receptor." Nat. Struct. Biol. **5**(11): 986-992.

Basu, S., A. A. Szewczak, et al. (2000). "Direct Detection of Monovalent Metal Ion Binding to a DNA G-quartet by ^{205}Tl NMR." J. Am. Chem. Soc. **122**: 3240-3241.

Bennett, M., A. Krah, et al. (2000). "A DNA-porphyrin minor-groove complex at atomic resolution: The structural consequences of porphyrin ruffling." Proceedings of the National Academy of Sciences of the United States of America **97**(17): 9476-9481.

Berman, H. M., W. Stallings, et al. (1979). "Molecular and Crystal Structure of an Intercalation Complex: Proflavine-Cytidylyl-(3'-5')-Gaunosine." Biopolymers **18**: 2405-2429.

Berman, H. M. and P. R. Young (1981). "The Interaction of Intercalating Drugs with Nucleic Acids." Ann. Rev. Biophys. Bioeng. **10**: 87-114.

Berman, H. M., C. Zardecki, et al. (1998). "The nucleic acid database: A resource for nucleic acid science." Acta Crystallogr. Sect. D-Biol. Crystallogr. **54**: 1095-1104.

Brunger, A. T. (1992). "Free R-Value - a Novel Statistical Quantity For Assessing the Accuracy of Crystal-Structures." Nature **355**(6359): 472-475.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr. Sect. D-Biol. Crystallogr. **54**: 905-921.

Canals, A., M. Purciolas, et al. (2005). "The anticancer agent ellipticine unwinds DNA by intercalative binding in an orientation parallel to base pairs." Acta Crystallographica Section D-Biological Crystallography **61**: 1009-1012.

Cesare Marincola, F., V. P. Denisov, et al. (2004). "Competitive Na(+) and Rb(+) binding in the minor groove of DNA." J. Am. Chem. Soc. **126**(21): 6739-50.

Chaires, J. B. (1997). "Energetics of drug-DNA interactions." Biopolymers **44**(3): 201-15.

Chaires, J. B., S. Satyanarayana, et al. (1996). "Parsing the Free Energy of Anthracycline Antibiotic Binding to DNA." Biochemistry **35**: 2047-2053.

Chiu, T. K. and R. E. Dickerson (2000). "1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium." J. Mol. Biol. **301**(4): 915-45.

Coll, M., C. A. Frederick, et al. (1987). "A BIFURCATED HYDROGEN-BONDED CONFORMATION IN THE D(A.T) BASE-PAIRS OF THE DNA DODECAMER D(CGCAAATTTGCG) AND ITS COMPLEX WITH DISTAMYCIN." Proceedings of the National Academy of Sciences of the United States of America **84**(23): 8385-8389.

Conn, G. L., A. G. Gittis, et al. (2002). "A compact RNA tertiary structure contains a buried backbone-K^+ complex." J. Mol. Biol. **318**(4): 963-673.

Correll, C. C., I. G. Wool, et al. (1999). "The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 A resolution." J. Mol. Biol. **292**(2): 275-87.

Crick, F. H. C. and B. S. Magdoff (1956). "The theory of isomorphous replacement for protein crystals." Acta Crystallogr. **9**: 901-908.

Dauter, Z. and D. A. Adamiak (2001). "Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy." Acta Crystallogr. D., Biol. Crystallogr. **57**(Pt 7): 990-995.

Davis, I. W., A. Leaver-Fay, et al. (2007). "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." Nucleic Acids Research **35**(Web Server issue): W375-83. Epub 2007 Apr 22.

DeLano, W. L. "The PyMOL Molecular Graphics System." DeLano Scientific LLC, San Carlos, CA, USA http://www.pymol.org.

Denisov, V. P. and B. Halle (2000). "Sequence-specific binding of counterions to B-DNA." Proc. Natl. Acad. Sci. U.S.A. **97**(2): 629-33.

Denny, W. A. (1989). "DNA-Intercalating Ligands as Anti-cancer Drugs: Prospects for Future Design." Anti-cancer Drug Design **4**: 241-263.

Doyle, D. A., J. Morais Cabral, et al. (1998). "The structure of the potassium channel: molecular basis of K^+ conduction and selectivity." Science **280**(5360): 69-77.

Egli, M. (2002). "DNA-cation interactions: Quo vadis?" Chemistry & Biology **9**(3): 277-286.

Egli, M., V. Tereshko, et al. (1998). "X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution." Biopolymers **48**(4): 234-252.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D-Biological Crystallography **60**: 2126-2132.

Farrugia, L. J. (1997). "Ortep-3 for Windows." J. Appl. Cryst. **30**: 565.

Fratini, A. V., M. L. Kopka, et al. (1982). "Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG." J. Biol. Chem. **257**(24): 14686-14707.

Gill, M. L., S. A. Strobel, et al. (2006). "Crystallization and characterization of the thallium form of the Oxytricha nova G-quadruplex." Nucleic Acids Research **34**(16): 4506-4514.

Glusker, J. P. (1994). Crystal structure analysis for chemists and biologists. N.Y., N.Y [i.e. New York, N.Y.] :, VCH.

Hamelberg, D., L. McFail-Isom, et al. (2000). "The Flexible Structure of DNA: Ion Dependence of Minor-Groove Structure and Dynamics." J. Am. Chem. Soc. **122**: 10513-10520.

Hamelberg, D., L. D. Williams, et al. (2002). "The Effect of Neutralized Phosphate Backbone on the Minor Groove Structure of B-DNA." Nucleic Acids Res. **30**: 3615-3623.

Hartmann, B., D. Piazzola, et al. (1993). "BI-BII transitions in B-DNA." Nucleic Acids Res. **21**(3): 561-8.

Hays, F. A., A. Teegarden, et al. (2005). "How sequence defines structure: a crystallographic map of DNA structure and conformation." Proc. Natl. Acad. Sci. U. S. A. **102**(20): 7157-62.

Heinemann, U. and C. Alings (1989). "Crystallographic Study of One Turn of G/C-Rich B-DNA." J. Mol. Biol. **210**(2): 369-81.

Heinemann, U. and M. Hahn (1992). "C-C-A-G-G-C-m5C-T-G-G. Helical fine structure, hydration, and comparison with C-C-A-G-G-C-C-T-G-G." J. Biol. Chem. **267**(11): 7332-41.

Hendrickson, W. A. (1985). "STEREOCHEMICALLY RESTRAINED REFINEMENT OF MACROMOLECULAR STRUCTURES." Methods in Enzymology **115**: 252-270.

Howerton, S. B., A. Nagpal, et al. (2003). "Surprising roles of electrostatic interactions in DNA-ligand complexes." Biopolymers **69**(1): 87-99.

Howerton, S. B., C. C. Sines, et al. (2001). "Locating Monovalent Cations in the Grooves of B-DNA." Biochemistry **40**(34): 10023-10031.

Hud, N. V. and J. Plavec (2003). "A unified model for the origin of DNA sequence-directed curvature." Biopolymers **69**(1): 144-58.

Hud, N. V. and M. Polak (2001). "DNA-cation interactions: The major and minor grooves are flexible ionophores." Curr. Opin. Struct. Biol. **11**(3): 293-301.

Hud, N. V., V. Sklenar, et al. (1999). "Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending." J. Mol. Biol. **286**(3): 651-60.

Jeffrey, G. A. and J. Mitra (1984). "3-CENTER (BIFURCATED) HYDROGEN-BONDING IN THE CRYSTAL-STRUCTURES OF AMINO-ACIDS." Journal of the American Chemical Society **106**(19): 5546-5553.

Johnson, C. K. (1976). ORTEP II, Oak Ridge National Labratory, Report RNL-5138. Oak Ridge, Tennessee.

Jones, T. A., J. Y. Zou, et al. (1991). "Improved methods for binding protein models in electron density maps and the location of errors in these models." Acta. Crystallogr. A. **47**(Pt 2): 110-119.

Kielkopf, C. L., S. Ding, et al. (2000). "Conformational flexibility of B-DNA at 0.74 angstrom resolution: d(CCAGTACTGG)(2)." Journal of Molecular Biology **296**(3): 787-801.

Klein, D. J., P. B. Moore, et al. (2004). "The contribution of metal ions to the structural stability of the large ribosomal subunit." RNA **10**(9): 1366-79.

Kleywegt, G. J., M. R. Harris, et al. (2004). "The Uppsala Electron-Density Server." Acta Crystallographica Section D-Biological Crystallography **60**: 2240-2249.

Kleywegt, G. J. and T. A. Jones (1995). "WHERE FREEDOM IS GIVEN, LIBERTIES ARE TAKEN." Structure **3**(6): 535-540.

Kleywegt, G. J. and T. A. Jones (1996). "xdlMAPMAN and xdlDATAMAN - Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets." Acta Crystallographica Section D-Biological Crystallography **52**: 826-828.

Kleywegt, G. J. and T. A. Jones (1998). "Databases in protein Crystallography." Acta Crystallographica Section D-Biological Crystallography **54**: 1119-1131.

Komeda, S., T. Moulaei, et al. (2006). "A third mode of DNA binding: phosphate clamps by a polynuclear platinum complex." J. Am. Chem. Soc. **128**(50): 16092-103.

Lavery, R. and H. Sklenar (1988). "The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids." J. Biomol. Struct. Dynam. **6**: 63-91.

Lavery, R. and H. Sklenar (1989). "Defining the Structure of Irregular Nucleic Acids: Conventions and Principles." J. Biomol. Struct. Dynam. **6**: 655-667.

Lerman, L. S. (1961). "Structural Considerations in the Interaction of DNA and Acridines." J. Mol. Biol. **3**: 18-30.

Lerman, L. S. (1964). "Acridine Mutagens and DNA Structure." J. Cell Comp. Physiol. **64 (suppl. 1)**: 1-18.

Loria, J. P. and T. Nowak (1998). "Conformational Changes in Yeast Pyruvate Kinase Studied by $^{205}Tl^{+}$ NMR." Biochemistry **37**(19): 6967-6974.

Lu, X. J. and W. K. Olson (2003). "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures." Nucleic Acids Res. **31**(17): 5108-21.

MacKerell, A. D. (2009). "Contribution of the Intrinsic Mechanical Energy of the Phosphodiester Linkage to the Relative Stability of the A, B-I, and B-II Forms of Duplex DNA." Journal of Physical Chemistry B **113**(10): 3235-3244.

Marcus, F. and M. M. Hosey (1980). "PURIFICATION AND PROPERTIES OF LIVER FRUCTOSE 1,6-BISPHOSPHATASE FROM C57BL-KSJ NORMAL AND DIABETIC MICE." Journal of Biological Chemistry **255**(6): 2481-2486.

McConnell, K. J. and D. L. Beveridge (2000). "DNA structure: What's in charge?" J. Mol. Biol. **304**(5): 803-20.

McFail-Isom, L., C. C. Sines, et al. (1999). "DNA structure: cations in charge?" Curr Opin Struct Biol **9**(3): 298-304.

McLachlan, A. D. (1982). "Rapid Comparison of Protein Structures." Acta Crystallogr **A38**: 871-873.

Mocci, F. and G. Saba (2003). "Molecular dynamics simulations of A center dot T-rich oligomers: Sequence-specific binding of Na+ in the minor groove of B-DNA." Biopolymers **68**(4): 471-485.

Morth, J. P., B. P. Pedersen, et al. (2007). "Crystal structure of the sodium-potassium pump." Nature. **450**(7172): 1043-9.

Moulaei, T., T. Maehigashi, et al. (2005). "Structure of B-DNA with cations tethered in the major groove." Biochemistry **44**(20): 7458-7468.

Muller, W. and D. M. Crothers (1975). "Interactions of Heteroaromatic Compounds with Nucleic Acids. 1. The Influence of Heteroatoms and Polarizability on the Base Specificity of Intercalating Ligands." Eur. J. Biochemistry **54**: 267-277.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr. **53**(Pt 3): 240-55.

Neidle, S. and Z. Abraham (1985). "Structural and Sequence-Dependent Aspects of Drug Intercalation into Nucleic Acids." CRC Rev. Biochemistry **17**: 73-121.

Neidle, S., L. H. Pearl, et al. (1987). "DNA Structure and Perturbation by Drug Binding." Biochemistry J. **243**: 1-13.

Nelson, H. C. M., J. T. Finch, et al. (1987). "The Structure of an Oligo(dA)•Oligo(dT) Tract and its Biological Implications." Nature **330**: 221-226.

Otwinowski, Z. and W. Minor (1997). Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Pt A. San Diego, Academic Press Inc. **276:** 307-326.

Otwinowski, Z. and W. Minor (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymol., Macromolecular Crystallography. J. Carter, C.W. and R. M. Sweet. New York, Academic Press. **276, Part A:** 307-326.

Pedersen, P. A., J. M. Nielsen, et al. (1998). "Contribution to Tl^+, K^+, and Na^+ binding of Asn776, Ser775, Thr774, Thr772, and Tyr771 in cytoplasmic part of fifth transmembrane segment in alpha-subunit of renal Na,K-ATPase." Biochemistry **37**(51): 17818-27.

Phillips, K., Z. Dauter, et al. (1997). "The crystal structure of a parallel-stranded guanine tetraplex at 0.95 angstrom resolution." Journal of Molecular Biology **273**(1): 171-182.

Pullman, A. and B. Pullman (1981). "MOLECULAR ELECTROSTATIC POTENTIAL OF THE NUCLEIC-ACIDS." Quarterly Reviews of Biophysics **14**(3): 289-380.

Rouzina, I. and V. A. Bloomfield (1998). "DNA bending by small, mobile multivalent cations." Biophys. J. **74**(6): 3152-3164.

Saenger, W. (1987). "Structure and Dynamics of Water Surrounding Biomolecules." Ann. Rev. Biophys. Chem. **16**: 93-114.

Schneider, B. and H. M. Berman (1995). "Hydration of the DNA bases is local." Biophysical Journal **69**(6): 2661-2669.

Schneider, B., K. Patel, et al. (1998). "Hydration of the phosphate group in double-helical DNA." Biophys. J. **75**(5): 2422-2434.

Sheldrick, G. M. (1997). "SHELX-97." SHELX-97, Gottingen University, Germany.

Sheldrick, G. M. (2008). "A short history of SHELX." Acta Crystallographica Section A **64**: 112-122.

Shieh, H.-S., H. M. Berman, et al. (1980). "The Structure of Drug-Deoxydinucleoside Phosphate Complex: Generalized Conformational Behaviour of Intercalation Complexes with RNA and DNA Fragments." Nucleic Acids Res. **8**: 85-97.

Shui, X., L. McFail-Isom, et al. (1998). "The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium." Biochemistry **37**: 8341-8355.

Shui, X., C. Sines, et al. (1998). "Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations." Biochemistry **37**: 16877-16887.

Sines, C. C., L. McFail-Isom, et al. (2000). "Cations mediate B-DNA conformational heterogeneity." J. Am. Chem. Soc. **122**: 11048-11056.

Soler-Lopez, M., L. Malinina, et al. (2000). "Solvent organization in an oligonucleotide crystal. The structure of d(GCGAATTCG)$_2$ at atomic resolution." J. Biol. Chem. **275**(30): 23034-44.

Sponer, J., J. Leszczynski, et al. (2001). "Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases." Biopolymers **61**(1): 3-31.

Suh, D. and J. B. Chaires (1995). "Criteria for the Mode of Binding of DNA Binding Agents." Bioorg. & Med. Chem. **3**: 732-728.

Tereshko, V., G. Minasov, et al. (1999). "The Dickerson Drew B-DNA dodecamer revisited at atomic resolution." J. Am. Chem. Soc. **121**: 470-471.

Tereshko, V., G. Minasov, et al. (1999). "A "hydrat-ion" spine in a B-DNA minor groove." J. Am. Chem. Soc. **121**(15): 3590-3595.

Tereshko, V., C. J. Wilds, et al. (2001). "Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments." Nucleic Acids Res. **29**(5): 1208-1215.

Villeret, V., S. Huang, et al. (1995). "Crystallographic Evidence for the Action of Potassium, Thallium, and Lithium ions on Fructose-1,6-bisphosphatase." Proc. Natl. Acad. Sci. U.S.A. **92**(19): 8916-8920.

Walsh, C. P. and G. L. Xu (2006). Cytosine methylation and DNA repair. DNA Methylation: Basic Mechanisms. Berlin, Springer-Verlag Berlin. **301:** 283-315.

Wang, A. H. J., G. J. Quigley, et al. (1979). "Molecular-Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution." Nature **282**(5740): 680-686.

Waring, M. J. (1981). "DNA Modification and Cancer." Ann. Rev. Biochemistry **50**: 159-192.

Williams, L. D. (2005). Between objectivity and whim: Nucleic acid structural biology. DNA Binders and Related Subjects. Berlin, Springer-Verlag Berlin. **253:** 77-88.

Williams, L. D., M. Egli, et al. (1992). DNA Intercalation: Helix Unwinding and Neighbor Exclusion. Structure & Function, Volume I: Nucleic Acids. R. H. Sarma and M. H. Sarma. Albany, Adenine Press**:** 107-125.

Williams, L. D. and L. J. Maher (2000). "Electrostatic Mechanisms of DNA Deformation." Annu. Rev. Biophys. Biomol. Struct. **29**: 497-521.

Wlodawer, A., W. Minor, et al. (2008). "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures." FEBS Journal **275**(1): 1-21.

Woods, K., L. McFail-Isom, et al. (2000). "Monovalent Cations Sequester within the A-tract Minor Groove of $[d(CGCGAATTCGCG)]_2$." J. Am. Chem. Soc. **122**: 1546-1547.

Young, M. A., B. Jayaram, et al. (1997). "Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets." J. Am. Chem. Soc. **119**: 59-69.

Zell, R. and H. J. Fritz (1987). "DNA MISMATCH-REPAIR IN ESCHERICHIA-COLI COUNTERACTING THE HYDROLYTIC DEAMINATION OF 5-METHYL-CYTOSINE RESIDUES." EMBO Journal **6**(6): 1809-1815.